2012

Organization of the Centromeric Satellite I Cluster and D21z1 Short Arm Junction Region of Human Chromosome 21

Riddhi V. Patel
Loyola University Chicago

LOYOLA UNIVERSITY CHICAGO

ORGANIZATION OF THE CENTROMERIC SATELLITE I CLUSTER AND D21Z1 SHORT ARM JUNCTION REGION OF HUMAN CHROMOSOME 21

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE GRADUATE SCHOOL

IN CANDIDACY FOR THE DEGREE OF

MASTER OF SCIENCE

PROGRAM IN BIOLOGY

BY

RIDDHI PATEL

CHICAGO, ILLINOIS

DECEMBER 2012

ACKNOWLEDGEMENTS

I would like to thank all of the people who made this thesis possible, starting with my wonderful professors in the Biology Department at Loyola University Chicago. Dr. Jeffrey Doering proved an excellent advisor, supporter and sounding board for me from the beginning of my time here at Loyola University Chicago Biology Department. I would like to thank my committee members- Dr. Michael Cummings, Dr. Howard Laten and Dr. Bryan F. Pickett for their constant guidance and helpful insights into my research. I would also like to thank Robert Ennesser for his technical guidance in the lab and William Ziccardi for his help with the formatting of my thesis. Lastly I would like to thank my family for their constant support throughout research and more during the writing process.

To my parents, Anuj and Dr.Jeffrey Doering

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Most Down syndrome (trisomy 21) cases result from nondisjunction in either meiosis I or meiosis II. In order to study nondisjunction, it is necessary to be able to accurately determine the parent of origin and stage of meiosis in which the aberrant chromosomal segregation occurred. This is currently done using polymorphic markers on the long arm of human chromosome 21 (HC21) which can undergo recombination relative to the centromere and thus give inaccurate results. I tested whether the satellite I cluster proximal to the centromeric alphoid cluster can serve as an HC21-specific centromeric marker. A previously isolated satellite I clone (pS21SatB) was sequenced and then tested for specificity to HC21. The results showed that this sequence is present on HC13 and HC21 and thus cannot serve as an HC21-specific centromeric marker. This study also characterized the heterogeneity of satI on HC21 and provided the first evidence of multiple satI families that are found on both HC13 and HC21.

Alphoid DNA is the only repetitive sequence present at the centromeres of all human chromosomes. The junction regions where alphoid DNA sequences transition into adjacent, non-repetitive DNA have not been fully characterized. The current map of HC21 has a gap in the p-arm alphoid (D21Z1) junction region adjacent to the centromeric satellite I sequence. My work filled this gap using YAC1F8 and completed the HC21 centromeric physical map, identified a novel Y-satellite I cluster and established clear linkage relations between D21Z1, satellite I and Y-Satellite I. The current unequal

crossover model for the evolution of alphoid cluster junction regions is based on the process of out-of-register recombination. The model predicts a gradual loss of sequence homogeneity as one moves towards the end of the alphoid array, but my work showed that the D21Z1 short arm junction region does not have this structure. Thus a different model is needed to explain the evolution of this region of HC21.

CHAPTER ONE

INTRODUCTION

Down syndrome (DS) arises due to an extra copy of chromosome 21 (trisomy 21). Apart from being the leading genetic cause of mental retardation, it also causes hypotonia, congenital heart defects and a higher than average risk of both Alzheimer-like dementia and leukemia (Hernandez and Fisher, 1996). In 95% of Down syndrome cases, trisomy results from nondisjunction in either meiosis I or meiosis II. The remaining 5% of cases arise from a Robertsonian translocation in which two nonhomologous acrocentric chromosomes fuse through their short arms.

In order to study nondisjunction, it is necessary to be able to accurately determine the parent of origin and stage of meiosis in which the aberrant chromosomal segregation occurred. This is currently done using polymorphic markers on the long arm of human chromosome 21 (HC21) (Kurnit et al., 1988; Chakravarti and Lynn, 1999; Lamb et al., 1996). Since these markers are located distal to the centromeric region, they can undergo recombination relative to the centromere. Thus, they can yield incorrect inferences about the parent of origin of the centromere and the stage of meiosis in which nondisjunction occurred. It is therefore necessary to find an HC21-specific centromeric marker that is also polymorphic (Laurent A-M. et al., 2003).

The human acrocentric chromosomes (13, 14, 15, 21 and 22) have their centromeres located much closer to one chromosome end, thus forming distinct short (p)

and long (q) arms. The short arms of all the acrocentric chromosomes are highly heterochromatic regions and consist of a variety of tandemly repetitive sequences (e.g., β satellite, and satellites I, II, and III) (Lee et al., 1997). These clusters of tandem repeats have a high degree of sequence similarity on all acrocentric p-arms. The difficulty in finding an HC21-specific marker is due to this sequence similarity in the heterochromatic p arms centromeres of the various acrocentric chromosomes, especially between HC13 & HC21 (Lo et al, 1999; Maratou et al., 2000). The sequence similarity is also thought to result in nonhomologous interactions between acrocentric chromosomes during meiosis, leading to mispairing and improper segregation. It has been hypothesized that these heterochromatic regions create a predisposition to the chromosomal nondisjunction events that cause Down syndrome (Choo, 1990; Vissel and Choo, 1991).

Satellite I is the most A-T rich tandemly repetitive DNA sequence (at least 70%) in the human genome, and is present on chromosomes 3, 4, 13, 14, 15, 21 and 22 (Meyne et al., 1993; Tagarro et al., 1993). The best characterized satellite I subfamily is pTR1-6 (Kalitsis et al., 1992), and when used to probe HC21 DNA, it detects at least two clusters of satellite I, one of which is at the distal end of the short arm and the other proximal to the centromeric region. The satellite I cluster most proximal to the D21Z1 centromeric alphoid cluster did not hybridize well to this probe, apparently because of a high degree of sequence divergence (Roy et al., 2000; Patel et al., 2005). This cluster is the only centromeric satellite I sequence found on any acrocentric chromosome and so I hypothesized that it would be a good candidate for an HC21-specific centromeric marker.

First goal was to test whether the satellite I cluster proximal to the D21Z1 alphoid cluster can serve as an HC21-specific centromeric marker. I completely sequence a previously-isolated clone (pS21SatB) from the proximal cluster and found that it contained 4.8kb of satI with a monomeric organization. The sequence was then BLAST analyzed to locate regions within it that had low similarity to all known satI sequences. Various sub fragments of this clone were tested for specificity to HC21 and all sub fragments showed presence on HC13 as well as HC21. Although this sequence can't serve as an HC21 specific centromeric marker, pS21SatB is the first satI clone obtained from chromosome 21 that is completely sequenced and studied in detail. This study also characterized the heterogeneity of satI on HC21 and provided the first evidence of multiple satI families that are found on both HC13 and HC21.

Alphoid DNA is the only repetitive sequence present at the centromeres of all human chromosomes (Manuelidis, 1978). It is a 171bp repeat that can be found in higher order repeat (HOR) structures (Willard & Waye 1987). Alphoid DNA is able to elicit centromeric activity in artificial chromosomes for a short time and thus is thought to be an important component of centromere function (Grimes and Cooke, 1998). D21Z1 is the major centromeric alphoid array on HC21 and is 1Mb in length consisting of homogenous 1.9kb higher-order repeats (HOR) (Ikeno et al., 1994). The regions where alphoid DNA sequences transition into adjacent, non-repetitive DNA have not been fully characterized. It is known that these junction regions have organizations quite different from the main alphoid array, i.e., they are monomeric alphoid DNA devoid of higher order repeat structures (Wevrick et al., 1992; Mashcova et al., 1996; Horvath et al.,

2000). It has been established that the junction regions are evolutionarily distinct from the main alphoid array (Bosovsky et al., 2004; Horvath et al., 2000; Scheuler et al., 2001). The current model for the evolution of these junction regions is based on the process of out-of-register recombination. It predicts that the two ends of the same alphoid array should have a gradual loss of sequence homogeneity at the same rate as one moves towards the ends of the array. (Willard et al 1993, 2006). The hypothesis that these two regions are indeed similarly organized has never been tested formally on any chromosome. Previous studies on D21Z1 show substantially different restriction patterns at the two ends of the cluster (So et al., 1997). While the HC21q-arm alphoid junction region has been studied (Bosovsky, 2004), the short arm junction region has never been mapped in detail. The current map of HC21 has a gap which consists of the p-arm alphoid junction region and satellite I sequence (Figure 6).

My second goal was to fill this gap and compare this region to the previously characterized q-arm junction region. The mapping of the gap was done with YAC1F8, which had previously been shown to contain the centromeric satI and D21Z1 clusters (Shukair et al., 2005). The YAC DNA was restricted and hybridized with probes for D21Z1, the centromeric satI cluster, and a Y-satI sequence to determine the physical linkage map between these clusters. My work has linked the HC21p-arm map, completing the HC21 centromere physical map, by identifying a novel Y-satI cluster and establishing clear linkage relations between D21Z1, satI and Y-SatI.

The characterization of the D21Z1 junction regions is also now complete. This map of HC21 can be used as a model for the structure of all acrocentric chromosomes

and particularly in understanding the function and evolution of the centromere in more detail. I have shown that the p-arm alphoid junction region does not show a gradual loss of sequence homogeneity. This is the first study where both end junctions of the same centromeric alphoid cluster have been studied in detail to determine the validity of out-of-register recombination model (unequal crossover model). The combined results of the current study and the work done by Bosovsky et al., (2004) suggest that HC21 alphoid junction regions contradict the prediction of the unequal crossover model. A different model is thus needed to explain the evolution of this region of HC21.

CHAPTER TWO

LITERATURE REVIEW

Centromeres

Originally identified as the primary constriction on metaphase chromosomes, the centromere of eukaryotes is a complex chromosomal substructure. This region is responsible for proper chromosomal segregation during mitosis and meiosis. The centromere functions in sister chromatid adhesion, kinetochore formation, pairing of homologous chromosomes and is involved in the control of gene expression (reviewed in Larin and Mejia, 2002; Henikoff et al., 2001; Koch, 2000). Along with elucidating aberrant chromosomal processes such as nondisjunction, the study of human centromeres helps in understanding the basic biology of chromosome function and cell division, and has broad ranging applications in human health and disease, including aneuploidy and birth defects, cell cycle regulation and cancer, and gene therapy.

The role of the centromere is conserved throughout evolution, though the sequences that accomplish centromere function in different organisms are not. Table 1 shows the major differences in the centromeric region of human, yeast and *Drosophila* (reviewed by Bjerling & Ekwall, 2002).

The centromeric heterochromatin of all human chromosomes is composed of clusters of tandemly repeated sequences (satellite DNA) that span megabases (Cooper et al., 1993). Different types of satellite sequences found at human centromeres include

satellites I, II, III, alpha and beta sequences. These satellite sequences are found in distinct arrays not intermixed with other classes of tandem repeats. The tandem clusters in a centromere can together span up to 5Mb. In addition to satellite repetitive sequences, interspersed repetitive elements are also found in centromere regions, including duplicons and retroelements like Lines-1 and Alu (Lee et al., 1997).

Centromere structure appears to be organized into different, separable domains in order to accomplish distinct centromere functions such as kinetochore assembly and sister chromatid adhesion. Though the DNA sequences that form a centromere in yeast, *Drosophila* and humans have revealed little conservation at the sequence level, their function is conserved throughout evolution. At the protein level, few centromere proteins are conserved in all these organisms and many are unique to different organisms. Structurally, the centromere is generally multilayered with a heterochromatin layer and a central core/inner plate region. It is becoming increasingly clear that the key factors for the assembly and function of the centromere structure are the specialized histones and modified histones which are present in the centromeric heterochromatin and in the chromatin of the central core. Thus, despite the differences in the DNA sequences and proteins that define a centromere, there is an overall structural similarity between centromeres of evolutionarily diverse eukaryotes (reviewed by Bjerling and Ekwall, 2002).

Down Syndrome

Trisomy 21, the chromosomal abnormality responsible for over 95% of Down syndrome (DS) cases, is the most common genetic cause of mental retardation in humans,

occurring in approximately 1 in 600–800 live births. Furthermore, these individuals represent a relatively small proportion of all trisomy 21 conceptions with over 80% of such conceptuses not surviving to term. Trisomy 21 fetuses account for approximately 1–2% of recognized spontaneous abortions. Thus, trisomy 21 is an important cause of human pregnancy failure, as well as a leading contributor to mental retardation (reviewed in Hassold and Sherman, 2000).

Down syndrome is associated with a characteristic facial appearance and poor muscle tone (hypotonia) in infancy. About 60% of DS patients are at an increased risk for congenital heart defects. Other symptoms include digestive problems such as gastroesophageal reflux or celiac disease, hearing loss, and patients are also at a 10 to30 fold increased risk of leukemia and Alzheimer-like disease (reviewed in Antonarakis et al, 2004).

Ninety five percent of individuals with DS have an extra chromosome 21 as a result of a meiotic nondisjunction error in the segregation of chromosomes 21 during the formation of gametes. Most of the remaining cases are due to translocations involving chromosome 21 and less than 1% of the cases are due to somatic mosaicism. (Hernandez and Fisher, 1996).

Non-disjunction

Nondisjunction refers to improper chromosomal segregation in meiosis I or II leading to disomic gametic cells, which give rise to anueploid offspring. Approximately 90% of the DS nondisjunction cases are maternal in origin, while the remaining 10% are paternal. In maternal cases, the majority (75%) of errors occur in MI while

nondisjunction occurs equally in MI and MII in paternal cases (Hassold and Sherman, 2000). Cases of maternally derived nondisjunction are positively correlated with increasing maternal age (Hassold and Sherman, 2000).

The percentage of trisomies among all clinically recognized pregnancies climbs from 2% for women <25 years of age to 35% for women >40 years of age (Hassold and Chiu 1985). Maternal age related nondisjunction could be due to: 1) an accumulation of toxic effects of the environment during the arrested state of the oocyte, 2) a degradation of meiotic machinery over time while in the arrested state, leading to a suboptimal resumption of MI and MII, 3) a change in ovarian functioning due to suboptimal hormonal signaling, or 4) degradation of the uterine environment (reviewed in Lamb et al., 2005).

Altered genetic recombination has been identified as a nondisjunction risk factor as well. In model organisms, absent or reduced levels of recombination, along with suboptimally placed recombinant events, increase the likelihood of malsegregation (reviewed in Hassold &Hunt 2001). In humans, no recombination or pericentromeric recombination levels less than half of normal increase the risk of meiosis I nondisjunction by 4.9 fold. Pericentromeric recombination rates more than twice normal lead to a 2.8 fold increased risk of meiosis II nondisjunction (reviewed in Hassold and Sherman, 2000; Lamb et al., 1996; Sherman et al., 1994).

Lamb et al. (1996) proposed a two hit model of nondisjunction in DS based on the contributions of altered recombination and increasing maternal age. The first step is formation of a susceptible homologue pair configuration. A meiosis I error occurs if there

is reduced recombination in the pericentromeric region leading to a loose association between the homologues and the possibility they may behave as nonhomologous chromosomes and segregate to the same cell. Then in meiosis II each homologue would act independently and one sister chromatid from each homologue would be found in the gamete. This would yield a disomic gamete with parental heterozygosity (one sister chromatid from each homologue). In a meiosis II error, a tight association between the homologues caused by increased recombination proximal to the centromere would prevent the homologues from segregating from one another in meiosis I. This association would than cause the homologues to segregate in meiosis II, rather than the sister chromatids segregating. This would lead to a disomic gamete with parental homozygosity for chromosome 21 and therefore be classified as an MII error. Thus it is hypothesized that all aberrant segregation occurs in meiosis I and is resolved in either of the two stages i.e meiosis I or II (reviewed in Hassold and Sherman, 2000; Lamb et al., 1996).

Increasing maternal age is the second component of Lamb's model. In most cases, the susceptible homologue configurations can still undergo proper meiotic division. However, as maternal age increases, the probability of aneuploidy occurring from one of these configurations increases. This may be due to the breakdown over time of meiotic proteins responsible for proper homologue segregation. This breakdown of key components would make it increasingly difficult for proper segregation of the susceptible configuration. Thus to summarize the 2 hit model: the first hit is the prenatal establishment of a susceptible tetrad and the second hit is disruption of a meiotic process

that increases the risk of nondisjunction of the susceptible configuration (Lamb et al., 1996).

To further understand the relationship between maternal age and altered recombination, Lamb et al. (2005) studied HC21 recombination patterns in 400 trisomy 21 cases of maternal meiosis I origin that were grouped by maternal age. There was no statistically significant association between age and overall rate of exchange. The placement of meiotic exchange, however, differed significantly among the age groups. Susceptible patterns (pericentromeric and telomeric exchanges) accounted for 34% of all exchanges among the youngest class of mothers but only 10% of those among the oldest class. The pattern of exchanges among the oldest age group mimicked the pattern observed among normally disjoining chromosomes 21. These results suggest that the greatest risk factor for nondisjunction among younger women is the presence of a susceptible exchange pattern and there is no association between age and overall exchange rates. In older women the meiotic machinery is less efficient and thus more error-prone. This makes even the non-susceptible exchange patterns susceptible to nondisjunction. Thus, in older women the proportion of nondisjunction due to normal exchange configurations increases as age-dependent risk factors exert their influence.

Acrocentric Chromosomes

HC21 is one of five acrocentric chromosomes (HCs 13, 14, 15, 21, and 22). Acrocentric chromosomes are characterized by having their centomere located close to one end of the chromosome, resulting in very short p-arms. Most of the genes found on the acrocentric chromosomes are located on the long arms (Eichler, et al., 2004). The

short arms are comprised mainly of heterochromatin. Heterochromatin is very tightly coiled and largely composed of repetitive DNA (Eichler, et al., 2004). Some of the major clusters of repetitive DNA found on acrocentric short arms include alphoid DNA, beta satellite, satellites I, II and III, 48 bp repeats and interspersed repeats like LINES and SINES. There is a great deal of similarity in the content and organization of these sequences on all the acrocentric p-arms. HC21 and HC13 have a particularly high degree of sequence similarity in all the major classes of repetitive sequences on their short arms and centromeric regions. It has been proposed that this similarity results in a greater tendency for acrocentric chromosomes to participate in non-homologous pairing, thus predisposing them to frequent non-disjunction (Choo, 1990).

The fact that most of the DNA is repetitive has made it difficult to create a physical map of the p-arms due to difficulty in sequencing these regions. In order to understand these regions better, a complete physical map of chromosome 21 might serve as a model for understanding the short arm of all other acrocentric chromosomes.

HC21p

The human genome sequencing project for HC21 found 127 known genes and 98 predicted genes and covered almost the entire long arm: 33,546,361 base pairs (Hattori, et al., 2000). The short arm and centromere of HC21 were left largely untouched. It is known that there is a cluster of ribosomal genes on the distal short arm and that the gene TPTE, which encode for a putative tryrosine phosphatase, is found on HC21p proximal to the centromere (Hattori, et al., 2000). Aside from this the heterochromatin composition,

organization, and functional significance of the centromere and short arm of HC21 are not well characterized.

It is known that satellites I and III, 48 bp satellite, beta and alpha satellites are all found in the centromeric and pericentromeric regions of HC21 (Lee, et al., 1997). The high degree of sequence similarities among acrocentric p-arms makes it difficult to find 21p-specific markers and hence the construction of genetic and physical map becomes challenging (Wang et al., 1999).

Various attempts have been made to create a detailed map of HC21 p. Wang et al. (1999) created a YAC- map of the HC21 p-arm (Figure 1). This map gave a basic marker framework for further structural study. A detailed physical map of HC21p, is being constructed by our lab (Figure 2). The structure appears to be a mosaic containing clusters of tandem repeats interspersed with low copy number repeats. The p-arm is ~20Mb in length, and many of the repetitive sequence families are at multiple locations both proximal and distal to the rDNA. There are several unexpected gene sequences, like TPTE and BAGE, in this otherwise dense heterochromatic region. The map is still incomplete, containing some gaps (Figure 2).

Satellite I

The first human centromeric DNAs to be identified were isolated from three genomic DNA fractions having slightly different buoyant densities in $CsSO_4$ gradients (reviewed in Lee et al 1997). These DNA fractions were referred to as classical satellites I, II, and III. Each DNA fraction consisted of a heterogeneous population of repetitive DNA sequences having similar buoyant densities. The sequence heterogeneity of these

DNA fractions prompted Prosser et al. (1986) to identify a single predominant family of simple repeats in each of these DNA fractions. These three simple sequence families were referred to as satellite families 1, 2, and 3, to indicate their enrichment in DNA fractions I, II, and III, respectively. The classical satellite I DNA fraction contained a 42 bp repeat (commonly referred to as satellite I), as well as a 2.5kb male-specific repeat (commonly referred to as Y-specific satellite I or DYZ2).

Satellite I (satI) is the most AT rich DNA fraction in the human genome (72.4%) and is the least abundant classical satellite. It consists of 42-bp repeats that are arranged as alternating 17-bp (ACATAAAATATCGAAAGT) and 25-bp (ACCCAAAATAGTAGTATTATATACTGT) repeat units. *In situ* hybridizations with specific DNA clones localized satellite I to the pericentromeric regions of chromosomes 3 and 4, and the pericentromeric regions and short arms of all the acrocentric chromosomes (Kalitsis et al., 1993; Meyne et al., 1993; Tagarro et al., 1993).

The first subfamily of human satellite I DNA to be characterized and cloned was pTRI-6, composed of 72 copies of a 42 bp monomer units organized as a 3.0kb higher order repeat. The overall conservation among the monomers is 85% and the AT content is 77%. pTRI-6 was hybridized to partial somatic cell hybrid panel, first at low stringencies and positive signals were observed on all acrocentric chromosomes. Under high-stringency conditions, however, hybridization signals were predominantly found on chromosomes 13 and 21, though chromosome 21 had a lower signal than that of chromosome 13. These results indicated that the pTRI-6 subfamily was also present on chromosome 21, though in significantly lower amounts or as only a partially-similar sequence (Kalitsis et

al, 1993). In order to find the precise location of sat I, alphoid and pTR1-6 probes were used to carry out double FISH analyses. This study revealed that on chromosomes 3 and 4, sat I is located on the q-arm adjacent to the alphoid domain. On acrocentric chromosomes, it's found at p13 loci. Along with the 13p13 locus on HC13, satI is also found at a 13p11.2 pericentromeric locus (Tagarro et al, 1994; Meyne et al, 1994). Kalitsis et al. (1993) used pTRI-6 as a probe to investigate a group of *de novo* t(14q21q) Robertsonian translocations under low stringency hybridization conditions. They concluded that the translocation event was accompanied by the loss of a significant portion, but not all of satellite I DNA, suggesting that satellite I could be playing a role in translocation.

DYZ2 (Y-specific satellite I)

Also present in the satellite I DNA fraction is DYZ2 (HC Y-specific satellite I). It is a 2.47kb tandemly repeated monomer unit containing one well-conserved Alu repeat sequence. The chromosome location of DYZ2 was found to be at the Yq12 heterochromatic region (Schmid et al, 1990). Initially this Y-specific satellite I was sequenced using 3 *HinfI* fragments the sizes of which were 775 bp, 820 bp and 875 bp. This satellite I is quite different from sat I found on other chromosomes because of the presence of a 775 bp fragment that is GC rich and has many restriction sites. It also has one well-conserved Alu repeat. The other 2 (850 bp &950 bp) fragments are AT rich but distinct in sequence from standard satellite I (Vincent et al., 1984). A 2.4-kb tandem repeat containing portions of the Y-specific SatI is found in regions associated with human chromosome 22q11.2 rearrangement disorders as well as in heterochromatic

regions, including Yq12 and the acrocentric chromosome short arms. This repeat contains an AT-rich sequence and an Alu repeat in addition to the YSatI. On HC21 the repeat is located in the pericentromeric p-arm (Babcock et al., 2007).

HC21 Satellite I

Previous work in our lab has indicated that there are at least two different subfamilies of satellite I on HC 21: the N6.4 subfamily (Musich & Dykes, 1986) and the pTRI-6 subfamily. Sequence comparison shows that these two subfamilies share only about 80% sequence identity on HC21 (Roy et al, 2000). The N6.4 satellite I like cluster is located on the short arm distal to the rDNA cluster and two pTRI-6-like sequence clusters are located on the short arm proximal to the rDNA and in the centromere (Figure 2, Roy et al., 2000).

There is no known satellite I cluster in the centromeric region of other acrocentric chromosomes (Trowell et al., 1993). The satellite I cluster (centromeric satI) most proximal to the D21Z1 alphoid cluster, is interspersed with regions that do not hybridize to pTRI-6 at moderate stringency, which suggests the presence of a variant satellite I sequence. This centromeric cluster is also polymorphic in size in the population, (Trowell et al., 1993; Roy et al., 2000), thus making this satellite I an excellent candidate for a HC21-specific centromeric marker.

In order to more fully characterize this HC21 centromeric satellite I, it was necessary to clone a small portion of this cluster. For this purpose, HC21 specific YAC831B6 (Wang et al., 1999), which includes D21Z1 and all of the centromeric satellite I cluster, was chosen. It was digested with *SalI* and the resulting fragments were

cloned into phagemid vector pBK-CMV. A random clone that hybridized with the satellite-I probe (pTR1-6) was subcloned into pSPORT. This sat-I clone is ~4.5kb in size and is called pS21SatB (Roy et al., 2000; Patel, 2005). Partial sequence analysis of this clone showed that there were regions that did not hybridize well with pTR1-6, which supported the hypothesis that this satI could serve as an HC21-specific marker.

Alphoid DNA

Alphoid DNA (α-satellite) is the only repetitive sequence that has been localized to the centromere of all human chromosomes (Lee et al 1997). This satellite DNA is composed of 171 bp repeats organized in tandem arrays. Independent 171 bp monomers of alpha satellite DNA generally exhibit substantial intermonomeric sequence divergence, on the order of 20-40%. However most of these monomers are organized in a hierarchical fashion, as constituent units of chromosome-specific higher order repeats (HORs). Related HORs exhibit less than 5% sequence divergence (Lee et al 1997). These alphoid arrays can be up to 4Mb in length in normal human centromeres (Maratou et al., 2000). The repetitive structure of alphoid DNA can be classified into two types of repeats (Ikeno et al., 1994): units composed of several monomers and organized as HORs (type-I alphoid repeat; α-I) and monomeric organization consisting of diverged alphoid monomer units (type-II alphoid repeat; α-II). Centromere components are primarily assembled on type-I alphoid sequences (Ikeno et al., 1994; Ando et al., 2002; Politi et al., 2002).

The kinetochore, a three-layered disk-shaped structure composed of a dense inner plate, a lucent middle domain, and a dense outer plate, is formed at the centromere at cell

division and is the attachment site for spindle microtubules. Centromeric proteins that are constitutive, such as CENP-A, -B, -C, -H and -I are present at the centromere throughout the cell cycle. CENP-A is related to histone H3 and is central to kinetochore assembly as it initiates centromere assembly. The kinetochore protein CENP-B has a 17 bp recognition sequence (CENP-B box) that appears at regular intervals in human centromeric alphoid DNA. It is speculated that the binding of CENP-B through the entire array of human alpha satellite may be important not only for achieving the centromere structure through nucleosomal spacing but it might also act as a bridge to connect domains of centromere chromatin that function in kinetochore assembly and those involved in heterochromatin-mediated sister chromatid adhesion (reviewed in Schueler and Sullivan, 2006). Recently it was reported that direct assembly of CENP-A and CENP-B in cells with synthetic alphoid DNA required functional CENP-B boxes. This is the first reported evidence of a functional molecular link between a centromere-specific DNA sequence and centromeric chromatin assembly in humans (Masumoto et al., 2002). Thus constitutive CENPs interact with alpha satellite DNA to create a specialized, higher order chromatin conformation that is important for centromere identity and kinetochore assembly.

The centromeric α-satellite cluster size ranges from 200kb to ~4Mb in the human genome. There is extensive variation in α-satellite array size between nonhomologous as well as homologous chromosomes. (Tyler-Smith and Willard, 1990). This leads to the suggestion that there is not a fixed amount of alphoid DNA necessary for centromeric function. In order to determine the role of alphoid DNA in centromere function Tyler-

Smith et al (1993) used chromosomal fragments of the Y chromosome containing alphoid DNA and found that several hundred kilobases of alphoid DNA and a small amount of adjacent sequence were sufficient for centromeric function. Alphoid clusters less than 200kb in length resulted in unstable chromosome maintenance at mitosis (Tyler-Smith, 1993). A study on the X chromosome centromere used two isochromosomes, both truncated within alphoid DNA. These chromosomal fragments stably segregated with the only common sequence being the DXZ1 alphoid array. This suggested that the alphoid array is sufficient to elicit centromeric function (Schueler et al, 2001).

Neocentromeres are chromosomal sequences that can serve as functional centromeres but do not contain alphoid DNA (reviewed in Larin and Lejia, 2002). Neocentromeres usually function only when a chromosome's normal centromere is deleted or rearranged. Chromosomes with neocentromeres segregate normally just like their alphoid DNA- containing counterparts. Thus, neocentromeres are able to behave functionally in all respects as if they are *bona fide* centromeres, which suggests that alphoid DNA is not always required for normal centromeric function. These observations thus favor the model that the basis for centromere activity in human chromosomes is epigenetic, rather than determined solely by primary sequence (Willard, 2001).

Arrays of tandemly repeated monomers may be evolutionary favored form of DNA sequence in and around centromeres because of their unique characteristic to maintain sequence homogeneity over long DNA segments, and, if necessary, during long time-periods. At the same time, satellite DNAs have enormous potential to change extremely rapidly in nucleotide sequence and/or in copy number (Mantovani et al, 2007).

The evolution of alphoid repetitive arrays can be explained by the out-of-register recombination model (Figure 3). When homologous chromosomes pair in meiosis I, the high degree of alphoid sequence similarity in the centromere allows misalignment of the arrays. Subsequent out-of-register recombinations can expand or decrease the number of tandem repeats in the array. Multiple rounds of such out-of-register recombination between HORs of α-satellite will lead to homogenization of sequences internal to the array. Since the edges of the array have both a decreasing likelihood of undergoing recombination because of the presence of similar sequence flanking only one end, as well as fewer possible unequal alignments, a gradual loss of homogeneity is expected towards the edges of the arrays (Figure 4; Smith, 1976; Willard et al., 2006).

HC21 Alphoid DNA: D21Z1 and α21-II

There are two distinct alphoid arrays on HC21: D21Z1 and α21-II which differ in their higher order organization. The D21Z1 locus contains a uniform array of an 11mer HOR that has frequent CENP-B motifs whereas α21-II contains diverged alphoid monomers with no HORs and few CENP-B boxes (Ikeno et al., 1994). The D21Z1 alphoid array is 1Mb in length with regular *EcoRI* restriction sites and the two edges of the D21Z1 array have different restriction patterns from one another (So et al., 1997), indicating different sequence organizations. This finding is unexpected, as the out-of-register recombination model predicts similar structures at both edges of repetitive arrays. The D21Z1 alphoid array is 99.7% identical to the alphoid array present on chromosome 13 and they are therefore indistinguishable from one another by current hybridization techniques (Maratou et al, 1999). α21-II present on the proximal short arm of human

chromosome 21 has numerous subfamilies of alphoid DNA. There are five major clusters of α21-II on HC21, known as M1, M2, M3, M4 and M5 (Figure 2). These clusters contain highly heterogenous sequences and extend nearly 5Mb from D21Z1 (Ziccardi et al., 2008). Within the alphoid sequences in this region are repetitive elements called low copy number repeats (LCNRs) and other interspersed repeats. α21-II may be involved in sister chromatid adhesion (He et al., 1998).

There is a much higher degree of variability in alphoid array sizes on chromosome 21 than other chromosomes. Interphase fluorescence *in situ* hybridization analysis (FISH) on alphoid arrays found very short arrays for 3.70% of HC21 compared to ≤0.12% for HC13&17 and 0% for other chromosomes. The same study also showed that short alphoid arrays of chromosome 21 were statistically more common in patients with Down syndrome (6.85%) than in control chromosome 21(3.70%) individuals. These short alphoid arrays ranged anywhere from 51kb to 184kb with a mean of 78kb (Lo et al, 1999).

Maratou et al. (2000) also studied variation in the HC21 alphoid DNA cluster size in relation to risk of nondisjunction. Based on pulsed field gel results, they observed differences in the size of alphoid arrays in 23 families tested. They concluded that the risk for nondisjunction is increased when the size of the alphoid array of one of the two HC21 homologues is particularly small. The size of these alphoid arrays is highly variable within the population, but this study found that mothers of DS children are more likely to have shorter than average alphoid arrays, suggesting that certain variations in the size and structure of these alphoid arrays may predispose a chromosome to nondisjunction. The

DNA samples in this study were digested with *BamH1*. This *BamHI* fragment includes satellite I as well as centromeric D21Z1 alphoid DNA (Figure 5). Thus when the size of the alphoid cluster was estimated from HC21 homologues of DS patients, it would include the size of the satellite I present in the *BamHI* fragment. Therefore the basis of the size variation could be the alphoid cluster, the satI cluster or a combination of the two.

Alphoid Junction Regions

Alphoid DNA clusters have been widely studied, but regions where the cluster transitions into adjacent, non-repetitive sequences are not yet well characterized. These alphoid junction regions are organized very differently from their adjacent main alphoid arrays. The alphoid DNA present in these regions seems to be monomeric alphoid devoid of higher order repeat structure. Thus the regularly repeating restriction sites and CENP-B boxes are not observed in these regions (Wevrick et al., 1992; Mashkova et al., 1996; Horvath et al., 2000). The junction regions have a high frequency of transposed elements such as Lines-1 (L1) and Alu within the alphoid sequence, relative to homogeneous alphoid arrays, and a higher frequency of these elements in the non-repetitive region (Mashkova et al., 1996; Schueler et al., 2001). The junction region of chromosome X alphoid showed a density of interspersed repetitive elements twice that of the genomic average (Schueler et al., 2001). These are all characteristics of heterogeneous alphoid DNA, and can be explained by the gradual loss of sequence homogeneity at the edges of repetitive arrays predicted by the out-of-register recombination model. Pericentromeric alphoid junction regions also frequently contain

large numbers of paralogous sequences. Paralogous regions are sequences derived from a common ancestor that have been duplicated and inserted throughout the genome. These may serve as a buffer between the transcriptionally inert heterochromatic centromere and adjacent euchromatic coding regions (Horvath et al., 2000, 2001).

The monomeric alphoid DNA present at these junction regions seems to be evolutionarily distinct from the associated main array. A phylogenetic study using 500 alphoid monomers from chromosome X revealed that the DXZ1 alphoid monomers formed an evolutionary clade distinct from the monomeric alphoid junction monomers. Using subfamilies of L1 retrotransposons known to be transpositionally active at different points in primate evolution, it was determined that DXZ1 was newly evolved or homogenized while the monomeric region was older (Schueler et al., 2001). The region also had sequence tag sites (STSs) that were specific to the X chromosome. Another study found that the monomeric alphoid present on chromosome 16 was no more closely related to the main array than it was to the alphoid arrays of any other chromosome (Horvath et al., 2000). Due to this apparent distinct evolutionary history, the alphoid junction regions on chromosome 21 may have centromeric sequences that have evolved independently of the main D21Z1 array.

q-arm Junction Region of HC21

The 115kb BAC clone 21B49A22 was the last long arm clone sequenced by the human genome project and this clone contains both the satellite sequences as well as unique sequences. Bozovsky et al. (2004) found that this clone contained 31.29kb (21B-α) of alphoid sequence (Figure 6). D21S190, the most pericentromeric long arm marker,

was found to be ~133kb away from the centromere proximal end of this BAC. Various Southern hybridizations were performed using alphoid probes at different stringencies to determine the linkage of this long arm sequence to the physical map of the centromere and to determine the organization of q-arm junction region (Figure 6). There is an apparent abrupt transition from non-repetitive DNA to monomeric 21B-α alphoid sequences and not a gradual loss of D21Z1 homogeneity. Phylogenetic studies of 21B-α showed that it was established and evolved independently of D21Z1 and that it is not simply the degenerate alphoid predicted by the unequal crossover model. The results are thus contradictory to predictions of the unequal crossover model of repetitive sequence evolution (Smith, 1978). The model predicts that both edges of the array will have a gradual loss of homogeneity at the same rate and therefore will have very similar structure. However, there is a difference in restriction pattern at the two ends of D21Z1 (Figure 5). Thus these results might have significant implications for Smith's model. The p-arm junction region organization of D21Z1 needs to be characterized in order to determine whether the out-of register recombination model is consistent with the evolution of the D21Z1 alphoid array.

Table 1. A survey of centromere DNA sequences and proteins in humans, Drosophila, fission yeast and budding yeast.

Centromere	Human	Drosophila	Fission yeast	Budding yeast
DNA sequences (references - see text)	Tandem arrays of 171 bp Monomer α-satellite repeats	Simple satellites and single complete transposable elements	A 15-kb central core with unique sequences (cnt and imr) flanked by 20-100 kb or (otr/K+L) repeats	Three conserved regions CDE-I TCACATGAT, CDE-II 80-90 bp >90% A+T, and CDE-III TGATTTCCGAA
Minimal DNA sequence requirement for function (references - see text)	<500 kb of DXZ1 α-satellites are sufficient for centromere function	420 kb is sufficient for centromere function on a Dp1187 minichromosome	7 kb of central core + 2 kb of the otr/K repeat sufficient for function	120 bp of DNA sufficient for function (9). Deletion of CDE-I or CDE-II reduces centromere function but point mutations in the central CCG in CDE-III completely inactivate the centromere

Table 1 – **Survey of centromeric sequences and proteins in various species.** (from Bjerling & Ekwall, 2002).

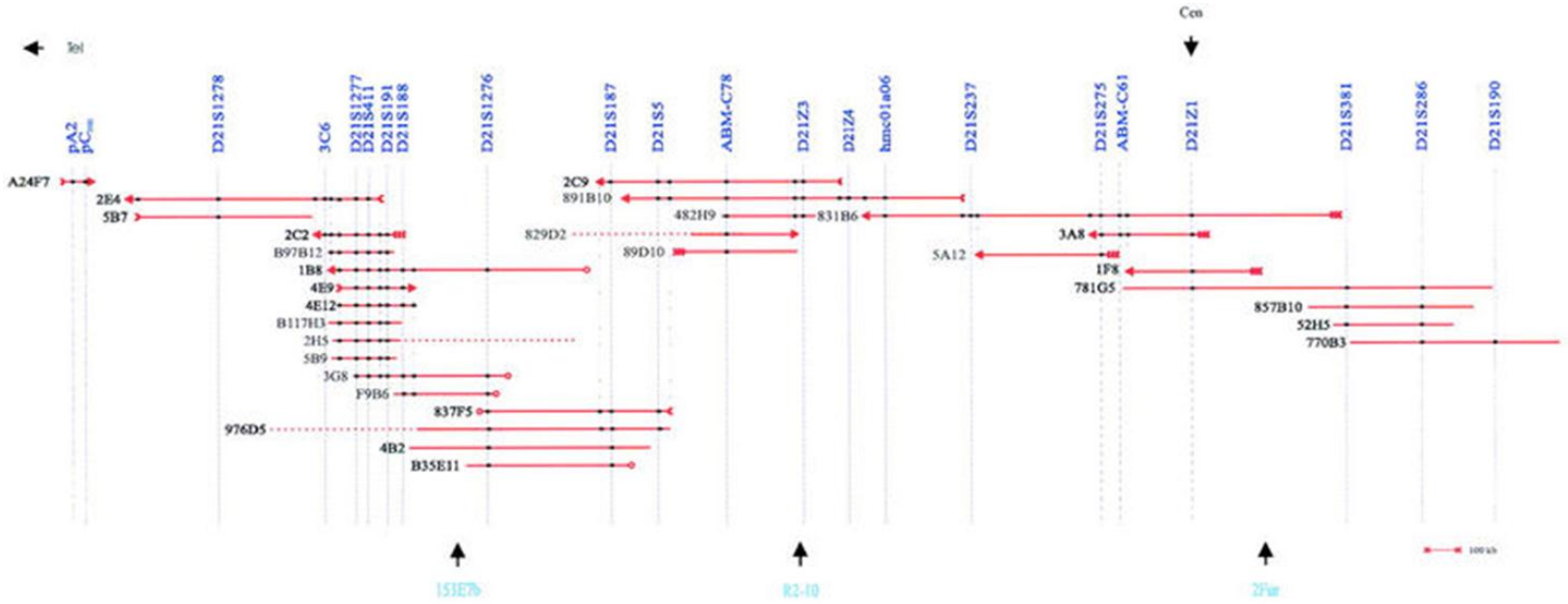


Figure 1. Chromosome 21p YAC contig map. The orientation of the YAC contig is based on the regional location of *D21Z1* and pC_{HB}. Horizontal lines represent YACs. Dotted lines indicate possible chimerical regions of the YACs. (Single arrowheads) Chromosome 21 YAC left (tails of arrows) and right (heads of arrows) ends; (triple tails of arrows) YAC end fragments homologous to α-satellite or satellite III repetitive sequences; (open circles) nonchromosome 21 YAC ends. (from Wang et al., 1999).

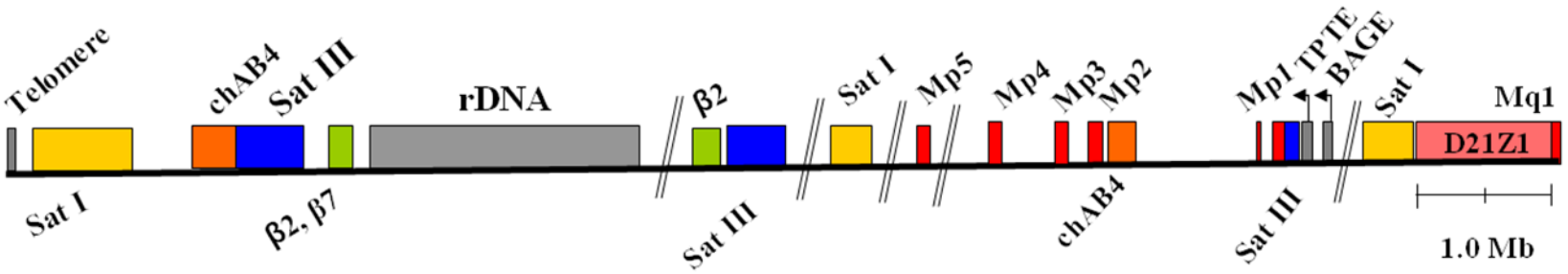


Figure 2. Map of HC21p. A scaled visualization of the current map of HC21p with key clusters of repetitive sequences noted. Hash marks represent gaps in the map. The distance of Mp5 from D21Z1 is notable as no other alphoid cluster has ever been found so far from the functional centromere.

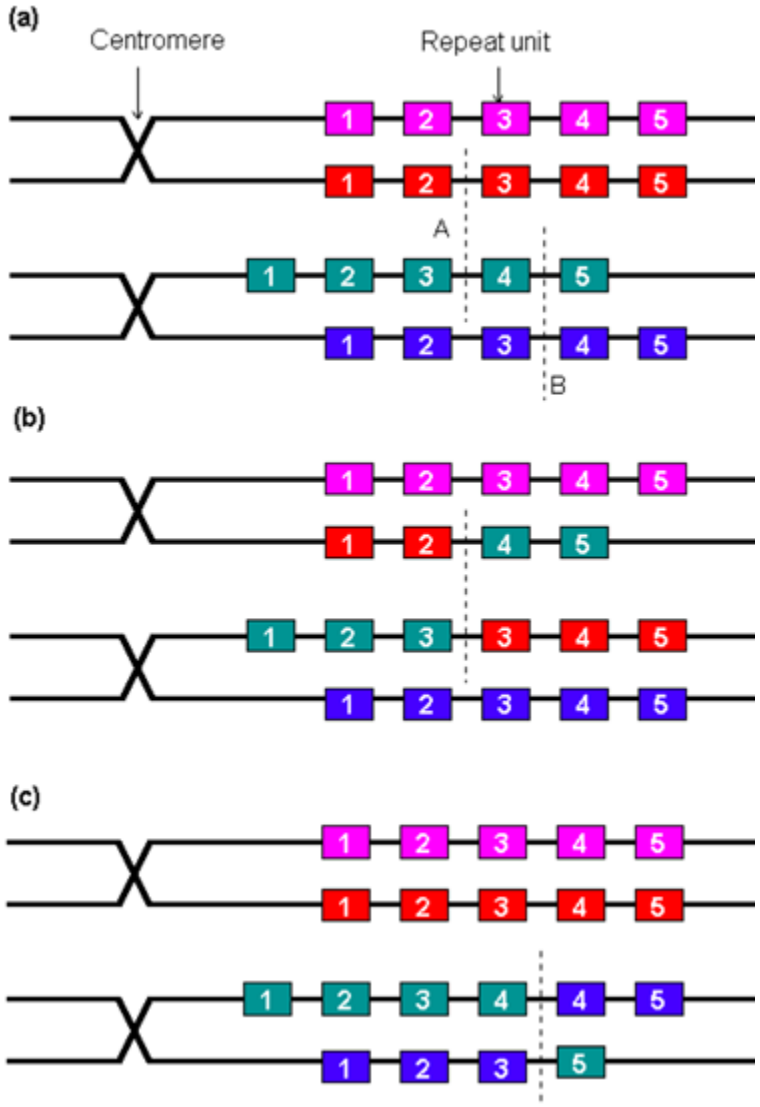


Figure 3. Representation of out-of-register recombination. Unequal crossover and sister chromatid exchange. (a) Two pairs of sister chromatids line up during meiosis. A repetitive region of one chromatid (the third one (green)) does not line up exactly with its corresponding region in other chromatid. (b) Strand breaks on the non-sister chromatid (along line A) resulting in unequal crossover, producing different number of repeat units in these chromatids. (c) Strand breaks on sister chromatids (along line B) also producing different repeats. In this case, it is called sister chromatid exchange.

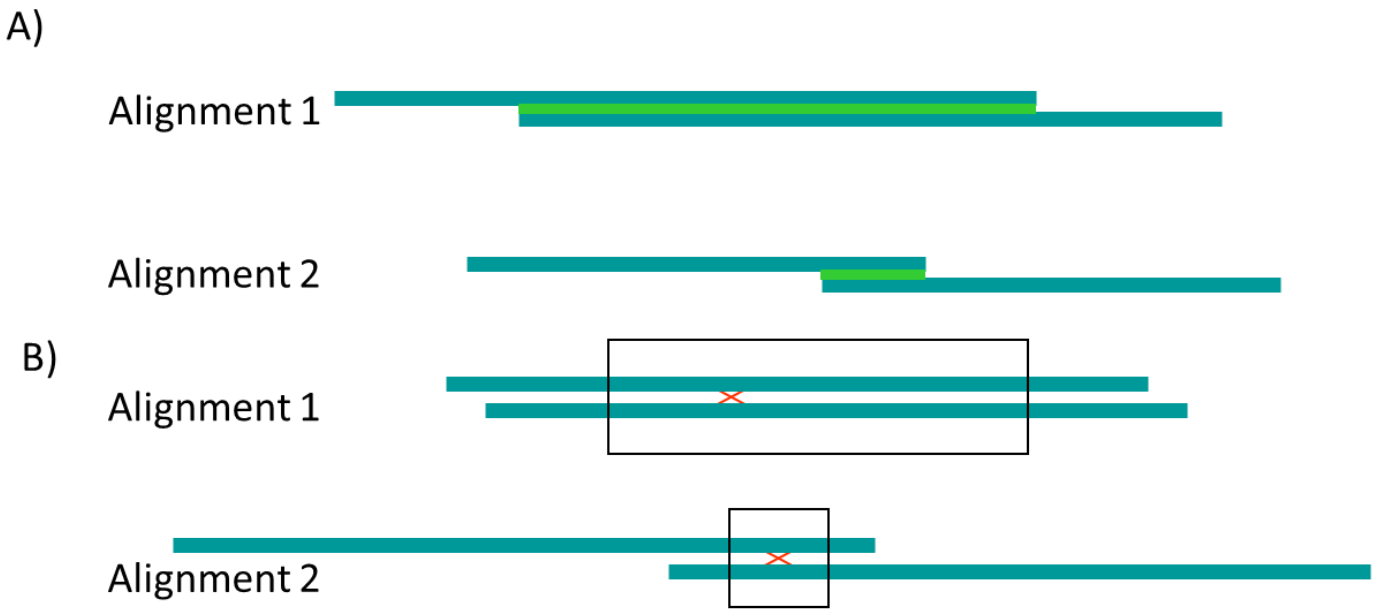


Figure 4. Events leading to decreased recombination in the distal array. A) Alignment 1 is more probable than alignment 2 due to the larger region of similar sequence (green region) B) In these alignments, the boxed region is more likely to undergo crossing over and therefore homogenization. Even in the event of alignment 2, the extreme edges have a low probability of undergoing recombination due to random distribution of recombination sites.

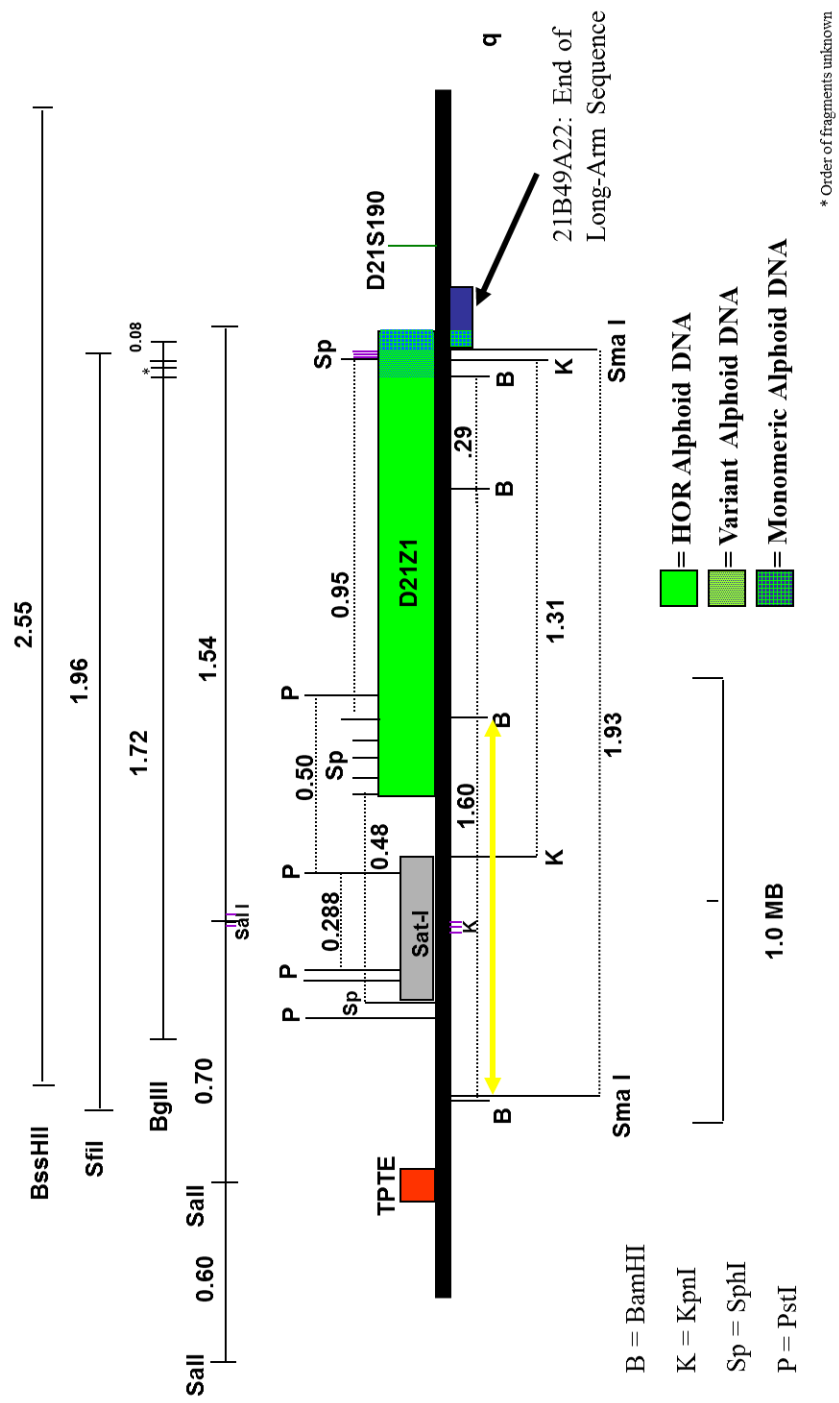


Figure 5. High resolution physical map of the centromere of HC21p-arm. The *BamHI* fragment shown in yellow colored line was used for the study of Maratou et al. (2000) which includes not only D21Z1(centromeric alphoid cluster) but also centromeric sat-I cluster.

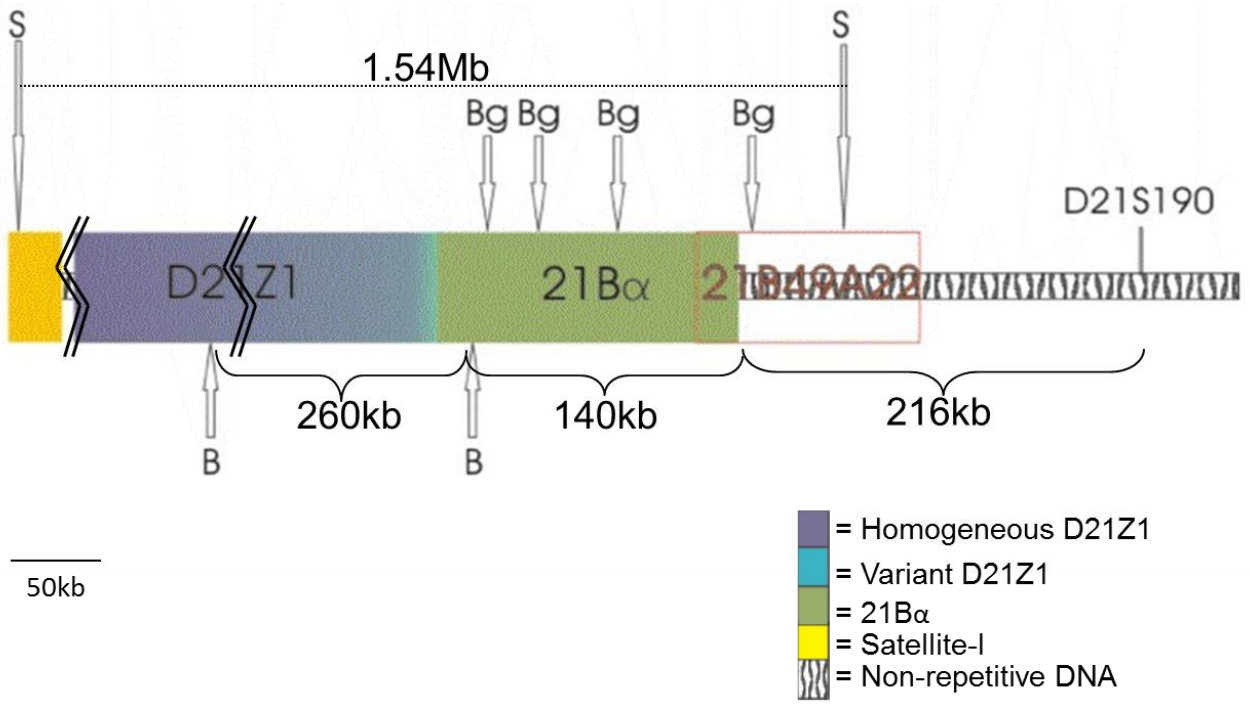


Figure 6. Physical map of the D21Z1 long arm junction region. Types of alphoid DNA present at the end of the D21Z1 array are shown. The location of BAC 21B49A22 is shown in red.

CHAPTER THREE

MATERIALS AND METHODS

Experimental Approach

My first goal was to characterize the organization of the centromeric satellite I cluster and determine if this satellite I is a good candidate for an HC21-specific centromeric marker. A satellite I-containing clone (ps21SatB) from HC21-specific YAC 831B6 (Wang et al., 1999) was previously obtained in our lab (Roy et al., 2000; Patel et al., 2005). This clone comes from the satellite I cluster proximal to the D21Z1 alphoid cluster. I completed the sequence of this ~4.5kb satellite I clone, and then did *in silico* sequence comparisons with already known satellite I sequences to see if this sequence is specific for HC21. Sequence specificity was also tested by hybridization between fragments of the satellite I clone and pTR1-6, a previously identified satellite I sequence (Kalitsis et al., 1992). The chromosome specificity of pS21SatB was also tested by hybridization to a hybrid cell (hamster-human) chromosomal panel in which each hybrid contains a different human acrocentric chromosome. This clone was also tested for polymorphism by hybridizing it to a panel of hybrid cell lines which contains HC21 from different individuals.

My second goal was to characterize the organization of the p-arm D21Z1 alphoid junction region of HC21 and compare it to the organization of the q-arm junction region. In order to fill the gap in the map and study the organization of the p-arm junction region,

I mapped that region using pulse field mapping of YAC1F8, which contains the short arm D21Z1 junction region of HC21. I selected restriction enzymes that rarely cut within pS21SatB and/or D21Z1 to develop a restriction map. Southern blotting was done using the D21Z1 alphoid probe (p11-4), and sat I probes pTR1-6 and 2.4kb SpeI clone (Babcock et al., 2007). The membrane containing 1F8 restriction digests was probed with one radioactive probe (e.g. p11-4). This probe was then stripped and the membrane was probed with the next probe (e.g. pTR1-6). Autoradiographs of the same blot exposed to different probes were superimposed to determine linkage. In order to determine the organization of the alphoid junction region, the probes were first hybridized at a normal stringency and thereafter at higher stringency. The monomeric sequences in the junction region that are heterogenous would not hybridize to p11-4 at higher stringency. Thus visual comparison of the sequence divergence restriction maps of the p and q arm junction regions showed how similar or different they are from each other.

Plasmid Purification

Plasmids containing the sequence of interest were purified using the Promega PureYield Plasmid Midiprep System, following the manufacturer's instructions.

Sequence Analysis

Samples to be sequenced were standardized to a concentration of 0.3-0.6µg/µl. A total of 10µl of the sample were sent to the sequencing facility at the University of Chicago. Primers were sent at a concentration of 2pmol/µl.

Programs for Sequence Determination and Analysis

Chromas: Chromas was the primary tool used for editing raw chromatograms downloaded from the University of Chicago website. It can be downloaded from http://www.technelysium.com.au

Lasergene: 1. SeqMan was used to assemble contigs and sequence chromatograms. Its primer walking feature was also used to design unique primers.

2. Primer Select was used to design sequence specific primers.

3. SeqBuilder was used to pick specific enzymes for YAC digests and also for finding enzymes for subcloning purpose. It was also used for predicting fragment sizes produced by particular restriction enzymes by inputing known insert and vector sequences.

4. EditSeq was used for editing final sequence and determining %A+T and %G+C content in specific clones for determining stringency of Southern blot post-hybridization washes.

5. DotPlot: this feature of Lasergene gave a graphical representation of the higher order repeating structure of any sequence.

NetPrimer: A free online tool that allowed checking various parameters of any given primer sequence, including its melting temperature, GC%, palindrome content and ability to form any hairpins, dimers and cross dimers. This helped determine if a particular sequence would be a good primer.

Link: http://www.premierbiosoft.com/netprimer/index.html.

Primer3Input: This program is an online tool at the University of Chicago sequencing facility website that was also used to design unique primers for pS21SatB sequencing. http://fokker.wi.mit.edu/primer3/input.htm

Tandem Repeat Finder: This program was used to determine the higher order repeating structure of pS21SatB. This program was run using the default alignment parameters for match, mismatch and indels of 2, 7 and 7 respectively and the minimum alignment score to report repeat was 50. Link: http://tandem.bu.edu/cgi-bin/trdb/trdb.exe

BLAST 2 sequences: This tool from the NCBI website was used to determine the % similarity between and within various sequences with the filter option turned off, and default conditions were: open gap penalty: 5, extension gap penalty:2, expect value 10.0, word size 11, reward for match:1 and penalty for mismatch: -2.

Restriction Digestions

All digests were done in the buffers and under conditions recommended by the manufacturers. For *KpnI* fragments of pS21SatB, 10µg pS21SatB was digested with 100 units of *KpnI* (Invitrogen), in a total volume of 80µl. For isolating the 1300 bp *NsiI* fragment from pUCSatB13, 20µg of plasmid were digested with 20 units of *NsiI* (Invitrogen) in a total volume of 100µl. For additional cloning, 1µg p21SatB was digested with *HindIII*(20 units), *MscI*(10 units), *PstI* (20 units, Invitrogen), in a total volume of 20µl. For cloning of KpnI fragments into pUC18, 0.2µg of pUC18 was digested with *KpnI* (40 units, Invitrogen), in a total volume of 20µl. Similar digestion was done for pGEM using *NsiI*(NEB). All the reactions were done at 37^0C for 2.5hours.

Mini-agarose Gel Electrophoresis

After the restriction endonuclease digestion, the DNA was run through a standard 1% ME agarose gel in a solution of 1X E buffer (made from 10X E Buffer: 0.4M Tris, 0.2M sodium acetate and 0.02M $Na_2EDTA.H_2O$ at pH7.8). Lambda/*HindIII* was run alongside DNA digest lanes as a standard for the purpose of calibrating DNA size. The gel was run at 80 volts for 2 hours, followed by 15 minutes of staining with 12µl of ethidium bromide(10mg/mL) in 350ml of 1X E buffer. The gel was then photographed for future records with Polaroid film.

DNA Isolation from Gels

The Gene Clean II system from QBioGene was used to purify specific DNA fragments from a gel. Four lanes of digested DNA were run on a gel. The gel was then cut in order to isolate the DNA fragment of choice which needed to be purified. The gel fragments were transferred to a pre-weighed microfuge tube. Depending on the weight of the fragment, three times as much NaI solution was added to the microfuge tube. This solution was then incubated in a 55˚C water bath in order to melt the agarose. When the agarose was fully dissolved, glassmilk corresponding to the amount of DNA as recommended by the protocol was added to the solution. The solution was then incubated at room temperature for 20 minutes so that the DNA bound to the glassmilk. The supernatant was discarded after centrifugation. The pellet was washed with NewWash twice and centrifuged each time. Sterile water corresponding to the amount of glassmilk added initially was added to the pellet. Resuspending the pellet in water causes the bound DNA to dissolve in water. After centrifugation the supernatant containing the DNA was

pipetted out into a new microfuge tube. The concentration of the purified DNA was estimated by running an aliquot on a gel alongside a defined amount of lambda/*HindIII* marker. Visual comparison of the purified fragment intensity with that of the known quantities in the lambda/*HindIII* fragments allowed for estimation of the concentration of the purified fragment.

Ligation Reactions for Subcloning

In the usual protocol for ligation of cohesive ends, the insert:vector molar ratio was at least 3:1. The reaction mixture contained total DNA amounts ranging from 0.01 to 0.1µg, T4 DNA ligase (0.1 units, Invitrogen) and ligation reaction buffer (Invitrogen) in a volume of 20µl. The reaction was incubated at 23-26^{0}C for 3 hours. A 1:5 dilution was used for transformation. Specific reactions varying in insert:vector ratio and units of ligase are as follows:

0.014µg of the 662 bp *KpnI* fragment (from pS21SatB) was ligated to 0.002µg of pUC18, 7:1 insert:vector ratio. All other volumes and conditions were according to the usual protocol. 0.063µg 1500 bp *KpnI* fragment (from pS21SatB) was ligated to 0.018µg of pUC18 4:1 insert:vector ratio. Additional units of ligase was added the next day and incubated at 14^{0}C overnight again to get better ligation results.

0.09µg 1300 bp *NsiI* fragment (from pUCSatB13) was ligated to 0.005µg pGEM (18:1 molar ratio) and a similar concentration and insert:vector ratio was used for p21SatB ligation to pUC18. An additional unit of ligase was added the next day and incubated at 14^{0}C overnight again to get better ligation results.

Transformation of Competent Cells

MAX Efficiency DH5α Competent Cells (Invitrogen) were transformed with the ligation reaction. The mixture of 100µl of cells and 1µl of 1:5 diluted ligation reaction was incubated on ice for 30 minutes. Sometimes undiluted ligation reaction mixture was also used. This was followed by a 45 second heat shock at 42^{o}C and 2 minutes on ice. S.O.C medium was added and the cells were allowed to recover at 37^{o}C with shaking at 225rpm for 1hour. X-gal(50µg/ml, 50µl) and IPTG(1 mM, 100µl) were added to the mixture(100-200µl) and it was then plated onto LB/amp plates and allowed to incubate overnight at 37^{o}C. White colonies were isolated, and re-plated since they contained inserts in the respective vectors.

In some cases dense plates contained unisolated white colonies intermixed with blue colonies. In order to isolate white colonies, fresh amp plates were plated with a mixture of 100µl of X-Gal and 50µl of IPTG. These plates were allowed to dry and then each white colony from the dense plates was replated using sterile toothpicks. If these white colonies were contaminated with blue colonies they were discarded. If there was no contamination, then white colonies were plated on fresh ampicillin plates from which CloneChecker analysis was performed.

CloneChecker System

White colonies from the ligation transformation reaction were selected for CloneChecker analysis (Invitrogen) to screen for colonies of bacteria that had picked up the cloned fragment of interest. The DNA was isolated from the colonies according to the protocol from the CloneChecker kit. A restriction digest was performed to excise the

cloned fragment of interest from the vector. Gel electrophoresis was carried out after restriction digestion to see which clones contained the desired fragment. Desired colonies were then grown up and plasmid DNA was purified for sequencing.

Making YAC plugs

Fifty ml of AHC medium were inoculated with 100µl of glycerol stock, that contained the desired YAC clone, and incubated with shaking at 30^{o}C for 1-2days. The number of cells was then counted using a hemocytometer. The appropriate volume of culture containing 1x10^{9}ml cells was placed in a 15ml Falcon tube and the cells were pelleted in a clinical centrifuge for 5 minutes and the supernatant was removed. The cells were resuspended twice in 2ml ET buffer (50 mM EDTA pH7.9, 1 mM Tris pH 8). The cells were then resuspended in 700µl of ET buffer and an equal volume of 2% LMP agarose was added with 25µl of 10mg/ml sterile Lyticase(Sigma). This mixture was then carefully added to the mold. The molds were then wrapped in parafilm and were allowed to solidify on ice for 2 hours. The plugs were gently pushed into a Falcon tube and washed twice for 24 hours each at 37 ^{o}C in 6ml of 10X ET buffer and 45µl of 10mg/ml Lyticase. The plugs were then incubated twice for 24 hours each at 50 ^{o}C in 10X ET buffer with 1%laurolyl sarkosyl and 1mg/ml proteinase K (Invitrogen) solution. The plugs were then transferred to ET buffer and stored at 4 ^{o}C.

Mapping YAC clones

YAC 1F8 (Wang et al., 1999) which contains some of the p-arm side of D21Z1 and the centromeric satellite I cluster was used for mapping. Each agarose plug contained 1.6µg of 1F8DNA. Before restriction digestion of plug DNA, EDTA storage solution was

removed by four washes with T_{10} E_1 (10 mM Tris, 1 mM EDTA pH 8.4) buffer of a half-hour each at 55°C. Single digests of the plugs were done using the enzymes *ApoI* (NEB), *BamHI* (Invitrogen), *BglII* (NEB), *HindIII* (Invitrogen), *HinfI* (Gibco/BRL), *PleI* (NEB), *PstI* (Promega/Invitrogen) *& SphI* (NEB). The plugs were incubated at 37°C for 15 hours in 5 volumes of 1x restriction buffer, 70 mM B-mercaptoethanol, 1 mM pCMB and 40 units of restriction enzyme. The reaction was stopped by two washes of T_{10} E_1 (pH 8.4) with 0.5% N-lauroyl sarkosyl and 1 mM pCMB at room temperature for 15 minutes. The plugs then had three room temperature washes with T_{10} E_1 (pH 8.4) and 1 mM pCMB for 20 minutes each.

The CHEF mapper (Bio-Rad) is a pulsed field gel electrophoresis system that can resolve DNA fragments from 3kb to 6Mb. The gel concentration and running buffers used depend on the size of fragments to be resolved. I used at least two different programs: One of the programs was able to resolve fragments ranging in size from 15-800kb. This program uses 1.0% pulsed field certified agarose with 0.5xTBE as running buffer. The running time is 23 hours 56 minutes at 6.0V/cm, at an angle of 120°. The other program resolves fragments ranging in size from 4-160kb. This program uses 1.0% pulsed field certified agarose with 0.5X TAE as running buffer. The run time is 9 hours 37 minutes at 6.0V/cm, at an angle of 120°. Following electrophoresis, the gels were blotted.

DNA Transfer and Hybridization

All gels were blotted using the alkaline Southern blotting technique (Reed and Mann, 1985). After the transfer of the DNA to a nylon membrane (Hybond), the

membranes were neutralized in 0.5M Tris pH 7.0 and 1.0M NaCl. Probes were radioactively labeled with ^{32}P using the Random Primers DNA Labeling System (Invitrogen).

The membranes were then incubated with shaking for 5 hours at 37^{o}C in a solution of 0.05M Tris pH 7.5, 50% formamide, 1.0M NaCl, 1% SDS and 10μg/mL of denatured salmon sperm DNA for pre-hybridization. The probe was denatured and added to the membrane in the prehybridization solution, and incubation proceeded overnight at 37^{o}C with shaking. Following hybridization, washes were done at the appropriate stringency. The first wash was done twice for 15 minutes each in 2xSSC at room temperature with shaking. The temperature and salt concentration for wash two varied depending on the stringency desired and was determined using the following formula: T_m=81.5 + 16.6 (log[Na^+]) + 0.41 (%G+C)-0.72 (% formamide)- (600/L). T_m is the melting temperature, M is the salt molarity, %G+C is the base composition of each probe and L is the length in base pairs of the hybrid formed (Sambrook et al., 1989). The second wash was done twice for 30 minutes each with shaking in the calculated concentration of SSC and 1%SDS at the appropriate temperature. The third wash was done twice for 30 minutes each at room temperature in the same SSC concentration as wash two. The membrane was exposed to x-ray film for varying periods of time with or without intensifying screens, as necessary, to provide clear images.

Hybrid Cell Lines

Chromosome specificity was tested using a monochromosomal hybrid cell panel that is composed of human/rodent cell lines containing each of the 24 human

chromosomes as the sole human component (Coriell Cell Repositories). Five- ten μg of each hybrid cell DNA were digested with 50 units of *KpnI* (NEB) in the manufacturer's recommended buffer for 2.5hrs at 37^{o}C and then run for 19hours on a vertical gel apparatus at 50volts with 1X running buffer. The DNA was then transferred to a membrane as described above.

CHAPTER FOUR

RESULTS

Sequencing pS21SatB

My first goal was to obtain the complete sequence of the satI clone pS21SatB. The insert size of this clone was estimated to be ~4.5kb (Patel et al., 2005). Universal primers were initially used to get sequence from both ends of the clone (Figure 7; Table 2; Patel et al., 2005) and primer walking was used to extend these sequences. As seen in Figure 7, sequence #1 was obtained by using the forward universal primer. From here on the sequences extending from sequence #1 will be referred to as "forward contig". Primers FP1(sequence #2) and FP2(sequence #3), designed from sequence #1, gave a total of 1626 bp of the forward contig. Sequence#5 was obtained by using reverse universal primer and the sequences extending from this sequence will be referred to as "reverse contig". Primer RP2 (sequence #4), designed from sequence #5, gave a total of 1324 bp of the reverse contig (Figure 7; Table 2). Several primers were designed from the ends of sequences 3 and 4 respectively to cover the gap (~1550 bp) and to get double coverage. These primers all yielded double sequencing ladders suggesting high internal sequence repetition. With primer walking having failed, transposon insertions were carried out in attempts to cover the sequence gap (Patel et al., 2005). Several transposon reactions were carried out but in most cases the transposon inserted at the vector-insert junction and yielded the same sequence that the universal primers gave (Figure 8; Table

3). One exception to this was transposon colony #5. Sequences 5FPedit (sequence #6) and 5RPedit(sequence #7) from the transposon reaction and primers designed from these sequences, 5FPU, 5FPL, and 5RP1R (sequences #8, 9 and 10 respectively) extended the forward contig to a size of 2.4kb. Given the ~4.5kb clone size the gap size was reduced to ~700 bp (Figure 8). No double coverage or extension was possible from the reverse contig as primers designed from this contig all gave double sequencing ladders. After several unsuccessful primer walking and transposon reaction attempts to fill the gap between the 2.4kb of the forward contig and the 1.3kb of the reverse contig of pS21SatB, a subcloning technique was utilized.

From the sequence available a *KpnI* digest of p21SatB was predicted to give ~6100, 1500, 662 and 252 bp fragments (Figure 9). Gel electrophoresis of the digest showed the predicted bands (Figure 10). To get double coverage for portions of the 1.3kb reverse contig and to cover the gap, the 662 bp and 1500 bp *KpnI* fragments were each gel extracted, subcloned into pUC18 and sequenced. Sequence analysis revealed that the size of the predicted ~1500fragment was actually 1800 bp instead. There was also a falsely predicted *KpnI* site on the reverse contig, i.e there was no *KpnI* site to give a 252 bp fragment. As a result of this the full sequence gap was not covered (Figure 11). This occurred because there was single coverage (from sequence#4, Figure 8) of poor quality and a double peak on the 2nd nucleotide (G) of the putative *KpnI* restriction site in the reverse contig. However since the gel showed a ~252 bp band it was assumed that this was the correct *KpnI* site. In fact the 252 bp fragment must be elsewhere. The false *KpnI* site was resolved with quadruple coverage of that region (Figure 12) showing that the

sequence was not GGTACC (*KpnI* restriction site) but GATACC (nucleotide position 3895) instead. The size of the reverse contig was thus 2436 bp.

Because the size of the clone was estimated to be ~4.5 kb and since there was 2.4 kb of forward contig and 2.4kb of reverse contig, an overlap of ~300 bp was expected (Figure 11). But the similarity of sequences in SeqMan was below 80% in the region where the forward and the reverse contigs overlapped, thus suggesting sequence contamination or false sequence alignment in the forward contig. Also instead of having a total of 3 *KpnI* restriction sites in pS21SatB (Figure 10) this alignment only gave 2 *KpnI* restriction sites (of the 662 bp fragment and 1800 fragments) for pS21SatB. The fact that there was no *KpnI* site in the forward contig (for the ~251 bp band) suggested some error in the forward contig sequence. The transposon-generated sequences were removed completely from the forward contig (sequences #6-12 Figure 8) to avoid any misalignments, and the size of the gap between forward and reverse contig was estimated to be ~300 bp (Figure 12; Table 4). This gap was covered by subcloning a 1300*NsiI* fragment into pGEM7Zf(-) and sequencing it using universal M13 primers and primer walking (Figures 13 and 14). Sequences #29-34 in Figure 15 show how the forward and the reverse contigs were joined once the 300 bp gap in the middle was covered. When pS21SatB clone was digested with *KpnI* it showed one extra fragment (size ~4361 bp) on the gel along with the ~6100, 1500, 662 and 252 bp fragments (Figure 10). To get rid of this contamination (~4361 bp band), I subcloned the satI insert from the original clone; p21SatB, into pUC18 and named it pUCSatB13. To verify the generation of a successful pure clone, it was digested with *KpnI* and universal primers M13 forward and M13

reverse were used to sequence the ends of the insert. (Figure 15 sequences #13,27, 36 and 37; Figure 16). The total size of this clone is 4778 bp and the completed contig with double coverage is shown in Figure 15. The sequences in Figure 15 are numbered and the corresponding primers sequences are listed in Table 5. The restriction map showing all the restriction enzymes used on pS21SatB is shown in Figure 17 and the complete sequence of pS21SatB is shown in Figure 18.

Sequence organization of pS21SatB

BLAST results with known satellite I sequences

In silico BLAST analysis of pS21SatB was carried out with all known satI sequences. The best matches over the longest distance are shown in Table 6. pTR1-6 is the satI sequence from HC13 (Kalitsis et al., 1993). It can be seen that there is high %match of pS21SatB to pTR1-6 especially near one end of the clone (Figure 19). There is also a 2.4kb region that has only 83% sequence similarity to pTR1-6. This suggests that there might be a putative HC21 centromere specific probe within this region. pN6.4 is a satellite I clone that is unassigned to any chromosome (Musich & Dykes, 1986) and it matches to pS21SatB at 82% over 605 bp with 22 gaps. Y-specific satellite I is a very different kind of satI compared to classical satI and it shows no sequence similarity to pS21SatB (Babcock et al., 2007). HC4 satI (Legrand, 1994) sequence is 83% similar to pS21SatB over 1059 bp with 35 gaps. BLAST2Seq results thus show that pS21SatB sequence has some regions within it that are not pTR1-6 like. These results also indicate that pS21SatB matches to all the other satI sequences at <85% and not over the entire length, suggesting that this sequence might serve as a chromosome-specific sequence.

Whole genome BLAST analysis

Whole genome (nt database) BLAST analysis was performed using pS21SatB sequence as the query sequence. The top 5 matches are listed in Table 7. The highest % similarity over the longest distance is listed to see how similar these 5 sequences are to pS21SatB over its entire length. Note that two are already-known satI sequences. Table 7 also shows the highest similarity of these 5 sequences to pS21SatB. This helps determine the regions within pS21SatB that might be putatively specific for HC21. pS21SatB matches to pTR1-6 at 96%. This match was over 629 bp with 2 gaps. pTR1-6 nucleotides # 2344-2970 matched nucleotides # 4146-4774 bp of pS21SatB. It is a significant match as the % similarity is very high and because there are only 2 gaps. It is important to notice that this match is over only 629 bp and not the entire length of pS21SatB. Chromosome 16 clone CTD-2506I5 is 167135 bp long and only 2.7% of this clone is satellite I. This clone has a match to pS21SatB at 81% over 499 bp. There were many other matches to this clone but all of them matched <80% to pS21SatB with >300 gaps. Chromosome 8 clone RP11-726G23 matches pS21SatB at 90%. The size of this clone is 192181 bp and it consists of 2.5% satI. This match was only over 54 bp with 3 gaps. All the other matches to this clone were <81% with >180 gaps. Chromosome 4 clone pSVH1-28 matches pS21SatB at 89%. The size of this clone is 1060bp and it is a satI clone completely. This match was over 164 bp with 3 gaps. All the other matches were <82% with >68 gaps. pS21SatB matches human middle repetitive DNA sequence L2HS6439 (a variant satI pN6.4) at 82%. This match was over 615 bp with 41 gaps. All other matches were <82% with >42gaps (Table 7). These results indicate that there are no

sequences in the database that are highly similar to pS21SatB (>85% and covering the entire sequence without any gaps), except pTR1-6 which matches to a specific region of pS21SatB at a very high similarity.

Tandem Repeats Finder analysis

In order to determine the monomer size and the number of monomers in this clone, Tandem Repeats Finder was run at default settings as described in Methods and Materials. This found 114 copies of a 41bp-repeating unit with a consensus of CCAAA-TATATATTATATACTGTACATAAAATATCAAAGTAC (Table 8). There is 1 bp difference in the consensus sequence of pTR1-6 and pS21SatB (Figure 20). pTR1-6 was shown to be a 42 bp repeating sequence (Kalitsis et al., 1992). The 42 bp consensus size of pTR1-6 was determined by lining up the 72 monomers that they sequenced. If a nucleotide was present in a particular position in greater than 50% of monomers it was considered as consensus. The degree of conservation among the monomers of pTR1-6 was around 85% (Kalitsis et al., 1992) whereas that for pS21SatB is 88% indicating that both sequences are very heterogeneous. As seen in Figure 20 when the reverse and complement of pTR1-6 is compared to the tandem repeats consensus for pS21SatB one A is missing in the pS21SatB consensus and thus it is 41 bp in size. Initially when Kalitsis et al. (1993) sequenced monomers of pTR1-6 that constitute the overall SatI consensus they found that both fragments had many variants and thus the repeat size for satI could be 41 bp or 42 bp. Figure 21 shows all 114 pS21SatB monomers lined up in tandem with the consensus sequence on the top.

Higher order repeating structure analysis

pTR1-6 is a 42 bp repeating unit arranged in a 3 kb higher order repeating(HOR) structure (Kalitsis et al., 1992). The pS21SatB sequence was divided into 11 equal segments which each consisted of 435 bp. Each one of these 11 segments was BLASTed against every other segment to see if there was any higher order repeating structure. If the degree of sequence matches between segments was above ~95%, it would be indicative of a higher order repeating structure. Table 9 shows the highest percent sequence similarity over the longest distance for each sequence comparison. The inter-sequence comparisons show that the sequence similarities are found in the range of 77-89%. Because the sequence similarities are low and highly variable among the 11 segments this suggests that pS21SatB is a monomeric satellite I sequence with no obvious higher order repeating structure. DotPlot analysis was then performed on the sequence to see if there is any evidence for an HOR over longer distances. The window size used was 41bp, the repeat unit for pS21SatB. Figure 22 shows the Dotplots of pS21SatB at various stringencies. The DotPlot result at 100% stringency shows that there is a clear direct repeat in pS21SatB but no HOR. The 1st of the direct repeats begins at bp#894 and ends at bp#2710. The 2nd repeat begins at bp#2700 and ends at bp#4470 (Figure 23). When the direct repeats were compared to each other using BLAST, the sequence similarity was found to be 96% over a length of ~1700 bp. A gradual increase in background is seen as the stringency is lowered to 95% and 90%. At 85% stringency very poorly conserved HORs are seen. The size of the HOR is ~500 bp and there are many breaks in it. Thus pS21SatB is monomeric satI but at low stringency very poorly conserved HORs do exist.

Identification of regions within pS21SatB putatively specific for HC21

Experimental evidence for the absence of higher order repeating structure in pS21SatB is shown in Figure 24. The acrocentric chromosomes were digested with *KpnI* and the blot was hybridized with pS21SatB at 97% stringency (0.15XSSC at 62.4^0C). HC21 does not show a 3kb band like that observed by Kalitsis et al. (1992) on HC13, indicating the absence of a higher order repeating structure in pS21SatB. This blot also shows that pS21SatB is not specific to HC21.

As the entire pS21SatB clone was not specific to HC21, the next goal was to test whether any of pS21SatB's internal fragments can serve as an HC21-specific centromeric sequence. In order to find a sequence that was different from the pTR1-6 sequence I used the eleven 435 bp fragments of pS21SatB to do a BLAST analysis against pTR1-6. The results are shown in Table 10. The criteria for selecting a putative probe sequence were: 1. The length had to be >1000 bp and 2. Its % similarity to pTR1-6 should be less than or equal to 85%. Thus I was looking for a probe sequence within pS21SatB with minimum similarity to pTR1-6 and maximum possible length. Fragments # 6 & 7 (nucleotides 2176-3045) have the lowest matches to pTR1-6 individually. Combined BLAST of segments 6 and 7(870 bp) against pTR1-6 showed 85% similarity, indicating that this sequence could be putatively specific for HC21. The combined % similarity is lower because the increase in size (from 435 to 870 bp), increased the # of gaps and hence reduced the % similarity. The next step was to find a restriction enzyme that would give the desired fragment, consisting of nucleotides 2176-3045, to use as a probe. The restriction enzyme *BstZ17I* creates a fragment overlapping this region (nucleotides 1821-

2997 bp) that has 83% similarity to pTR1-6. The location of this 1176 bp BstZ17I sequence is shown in Figure 25.

Testing 1176 bp BstZ17I fragment for chromosome specificity

A blot containing DNAs from each of the acrocentric chromosomes was used to test this fragment for chromosome specificity. The 1176 bp BstZ17I probe was gel-extracted and hybridized at 95% stringency. This stringency was used since the probe sequence and pTR1-6 sequences are only 83% similar. The intensity of hybridization to HC13 was much stronger compared to HC21 (very faint, Figure 26) suggesting that this probe was not specific to this particular HC21 copy. It also suggested that though there are no other acrocentric chromosomes that have a centromeric satI cluster, there might be pS21SatB-like sequences present elsewhere on HC13, that are not yet characterized. In order to further test whether this probe is specific to HC21, another copy of HC21 (Coriell Repository cell line # NA10323) from a different individual was used. Similar results (Figure 27) were obtained, suggesting this probe sequence is not specific for HC21. In order to understand the contradiction between sequence analysis results and blot results, I looked closely at the BLAST results of 1176BstZ17I against pTR1-6 and observed that there was a region within the 1176 bp sequence (bp position 2699-2896) that was ~190 bp long and was a 98-100% match to pTR1-6 (bp position 1411-1214, Figure 28). This was the likely reason why the 1176BstZ17I probe did not show specificity for HC21.

Testing 380bpBstZ17I fragment for chromosome specificity

Since the BLAST analysis and previous Southern blot results suggested that there were regions in SatI that did not hybridize well to pTR1-6, the criteria for picking the probe were modified. The length was significantly reduced to 300-700 bp and the % similarity to pTR1-6 was reduced to 95%. I looked for restriction enzymes that would give fragments within this range that did not have a stretch of nucleotides matching pTR1-6. Two such fragments were found: a 380 bp *BstZ17I* and a 736 bp *MseI* fragment (Figures 29 and 30). The 380 bp Bst*Z17I* fragment matched to pTR1-6 with 93% maximum similarity over its entire length, while the 736*MseI* fragment matched to pTR1-6 at 87% maximum similarity over its entire length.

To test how different these 2 fragments are from pTR1-6 in a hybridization assay a reverse Southern blot was made. This membrane contained all the *BstZ17I* fragments of pS21SatB in one lane and all the *MseI* fragments of the 1800 bp KpnI fragment in a second lane. This membrane was hybridized to pTR1-6 first at 62.3^0C with 0.15X SSC (>98% stringency). This stringency was chosen because both the sequence analysis of the 380 bp BstZ17I fragment and the 736 bp MseI fragment showed that they matched to pTR1-6 at or less than 93%similarity. Figure 31a shows that even at this stringency both the bands did hybridize to pTR1-6 but at a very low intensity compared to all the other bands on the membrane. To see if the hybridization intensity would reduce at a higher stringency, washes were done on the same blot at 65^0C with 0.15X SSC. Figure 31b shows that the 736 bp MseI band still showed a relatively strong hybridization signal compared to the 380 bp BstZ17I fragment, which had a very faint signal. These results

show that the similarity between pTR1-6 and the 380 bp BstZ17I fragment is much less than the similarity between pTR1-6 and the 736MseI fragment. Thus the 380 bp BstZ17I fragment was chosen to test for chromosome specificity.

The 380 bp BstZ17I fragment was hybridized to an acrocentric chromosome-containing hybrid cell panel first at normal stringency (2X SSC, 60^0C) to see to which chromosomes this probe hybridizes and at what intensity. Figure 32 shows very high hybridization intensity on HC21 compared to HC13, which suggested that this sequence might be specific for HC21 at higher stringency. The stringency washes were then carried out at 60^0C with 0.15X SSC (96% stringency). The results show (Figure 33) that even at this high stringency a strong band on HC13 is seen along with faint bands on HC21. At an even higher stringency of 62.5^0C with 0.15XSSC (99% stringency), HC 13 bands were still visible (Figure 34). This suggests that there are pS21SatB-like sequences on HC13. These sequences are likely present somewhere else on HC13 other than the centromeric region because previous work by Trowell et al. (1993) indicate that no other acrocentric chromosome other than HC21 has satI in the centromeric region.

Presence of Y-satellite I on HC21 and its linkage to centromeric D21Z1 and SatI clusters

In the process of determining the linkage between the centromeric clusters of satI and alphoid DNA on HC21 a publication by De Sario et al. (2003) was analyzed. This paper suggested that contig containing the centromeric end of HC21p (AL163201) had been completely annotated over its final 0.3 Mb. At the centromeric end of the contig, they found satI sequence that was about 0.6 kb long and matched to GenBank Accession

Number X00470 (HSATI- human satellite I sequence, Figure 35) at 80% nucleotide similarity. My closer look at this sequence showed that the size of this sequence is actually 0.76 kb and the identity to X00470 is 82%. Analyzing the sequence of this region further showed there was actually a 1.2 kb domain following the 0.76 kb region that was not annotated or characterized. This region had one AluSc element followed by 1 kb of simple TA repeats and 200 bp that did not match satI or the X00470 sequence. Preliminary BLAST sequence analysis showed that this ~2.0 kb region at the end of the contig had YsatI sequence. Y-satI was initially sequenced in part by Prosser et al. (1984) and Fenton et al. (1992, unpublished data). The 3 *HinfI* fragment sequences that are part of the Y-satI are shown in Figure 36. Prosser et al. (1984) sequenced 775 bp of X00470 in addition to 130 bp of M95571. They also sequenced 320 bp of X00490 and 130 bp of the end part of M95571. The rest of sequence of M95571 comes from the Fenton et al. (1992) group. Since the sequence comes from 2 different sources and has many unassigned and ambiguous nucleotides it was difficult to compare this Y-satI against Y-satI on HC21. Babcock et al. (2007) sequenced a 2.4 kb *SpeI* clone of Y-satI from HCY and this sequence is of much better quality. Table 11 shows the BLAST2Seq results of Y-specific satI on HC21 against the 2.4 kb *SpeI* clone, and the matches are in the range of 76-90%. Based on these results, a map of the organization of Y-satI on HC21 is shown with reference to the 2.4 kb *SpeI* clone (Figure 37). Thus this 2.0 kb sequence at the end of AL163201 contig is actually a Y-SatI sequence on HC21 that is not very well-conserved.

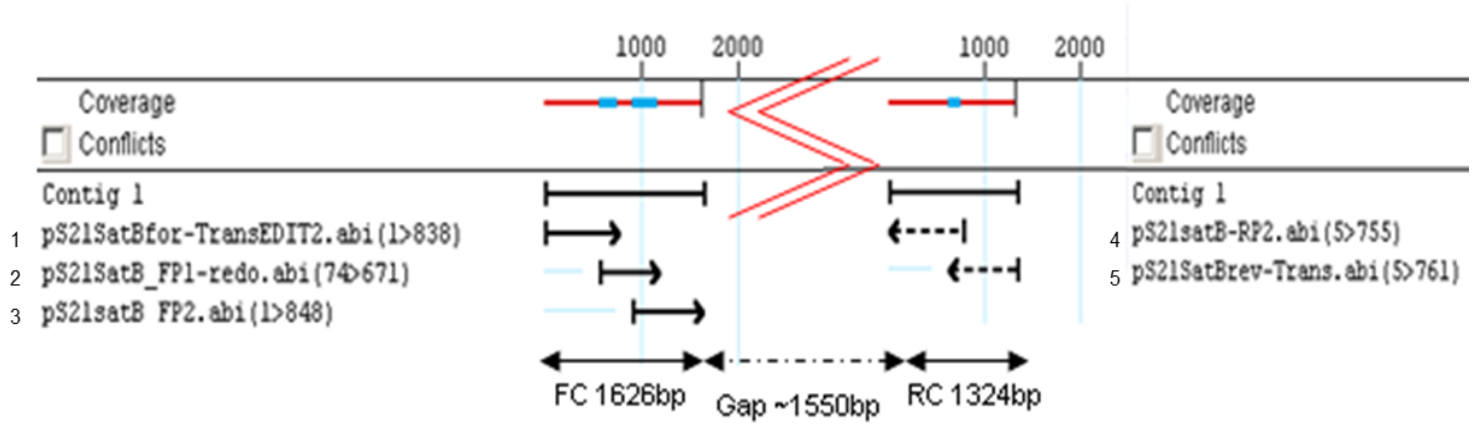
Determining linkage between the centromeric clusters

In order to determine linkage between D21Z1, satI and Y-satI, YAC1F8 (Figure 38) was digested with various restriction enzymes, and then hybridized first to an alphoid probe (p11-4) and subsequently to satellite I (pTR1-6) and Y-specific satellite I (2.4 kb *SpeI)* sequences respectively at normal stringency. Figure 39 shows the blot results for all three hybridizations. The size calibration of the bands for each probe is shown in Table 12. Based on these results, the linkage relationships and updated centromeric map are shown in Figure 40. The centromeric map has D21Z1, satI and Y-specific satI linkage relationships characterized. The order of the fragments is unknown except for the 58 kb and 115 kb BamHI fragments. The D21Z1 and satI junction region is 58 kb in size. This shows that the D21Z1 and satI clusters are much closer than earlier estimates. The gap between D21Z1 and satI was previously determined to be 76 kb (Bozovsky et al., 2004), but this map now shows the distance to be at most ~50 kb.

Determining the short arm junction region of D21Z1

To test the level of alphoid sequence heterogeneity at the p-arm end of the D21Z1 cluster, a membrane containing YAC1F8 digested with *BamHI, HindIII and SphI* was hybridized first to p11-4 at normal stringency (70%, 2XSSC, 60^0C). The same membrane was then hybridized at 75%(1XSSC, 60.5^0C), 85%(0.4XSSC, 63.9^0C) and 95%(0.1XSSC, 63.9^0C) stringencies. The blot results at all these stringencies are shown in Figure 41. As an internal control, the intensity of the >200 kb band, representing the main D21Z1 cluster, was maintained after each stringency wash. The 58 kb BamHI

fragment is indicative of the D21Z1 and satI junction region. There is no change in the relative intensity of this band at any of the stringencies suggesting that there is no significant loss in homogeneity within the D21Z1 alphoid array at the p-arm end. All the other bands except the >200 kb band are partial digestion bands because the size of YAC1F8 is 400 kb and the total of all BamHI fragments is ~600 kb. These results are consistent with no increase in the alphoid sequence heterogeneity at the end of the cluster, but rather an abrupt end of D21Z1 on the p-arm end. This result does not support the out-of-register recombination model, which predicts a gradual decline in homogeneity at the ends of a tandemly repetitive sequence cluster.



Forward Contig (FC): 1626bp
Reverse Contig(RC): 1324bp
Gap ~1550bp
Total size of the clone ~4.5kb

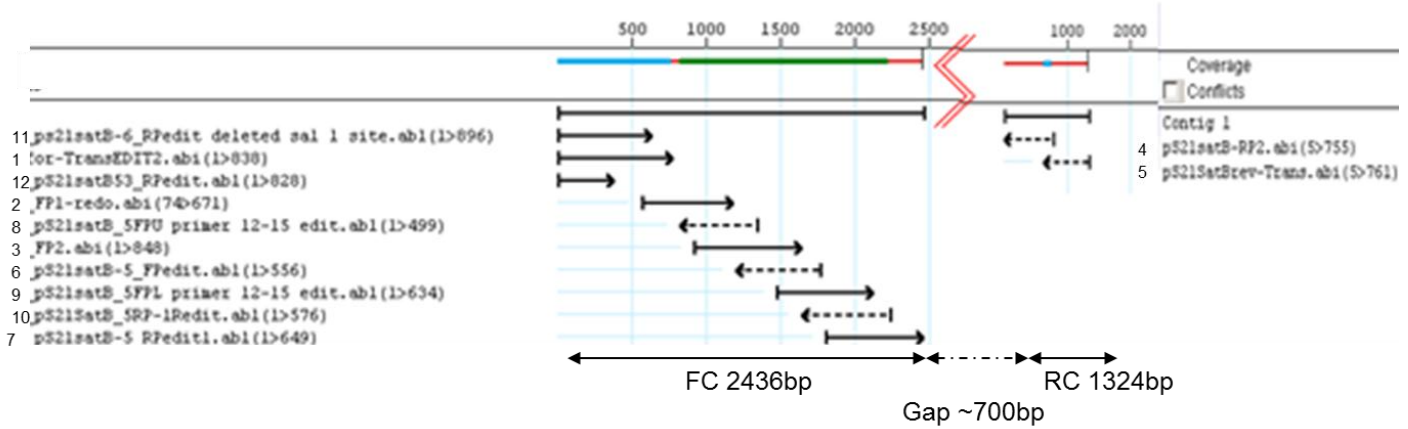
Figure 7: Initial sequence contig of pS21SatB. Forward contig is 1626bp and reverse contig is 1324bp. This sequence was obtained using the universal primers:

1. pS21SatBfor-TransEDIT2
5. pS21SatBrev-Trans

Primer walking sequences are sequences # 2,3 and 4

Sequence #	Position on the contig	Primer Sequence
1	1	universal primer (forward)
2	493	TATCAATGTACTCCAACTAT
3	909	ATCAAAGGACCCAAAATAAAC
4	4211	TATATTGGATGTGCCTTCA
5	4778	universal primer (reverse)

Table 2: List of primers used to build contig shown in Figure 7. The sequence # corresponds to the sequences labeled numerically in Figure 7.



Forward Contig (FC): 2436bp
Reverse Contig(RC): 1324bp
Gap ~700bp
Total size of the clone ~4.5kb

Figure 8: Sequence contig of pS21SatB generated using transposon reactions. Forward contig is 2436bp and reverse contig is 1324bp. The extension of the forward contig was obtained using transposon colony #5. Colony #6 and #53(sequences #11 and 12) gave sequences that overlapped with the universal forward primer sequence.

Sequence #	Position on the contig	Primer Sequence
1	1	universal primer (forward)
2	493	TATCAATGTACTCCAACTAT
3	909	ATCAAAGGACCCAAAATAAAC
4	4211	TATATTGGATGTGCCTTCA
5	4778	universal primer (reverse)
6	1187	transposon sequence from colony #5
7	1788	transposon sequence from colony #5
8	820	TATATTTGGGCGATTTA
9	1463	TCAAAGTACCCAAAGTATGC
10	1634	ATAATATACAGTTTGAGT
11	1	transposon sequence from colony #6
12	1	transposon sequence from colony #53

Table 3: List of primers used to build contig shown in Figure 8. The sequence # corresponds to the sequences labeled numerically in Figure 8.

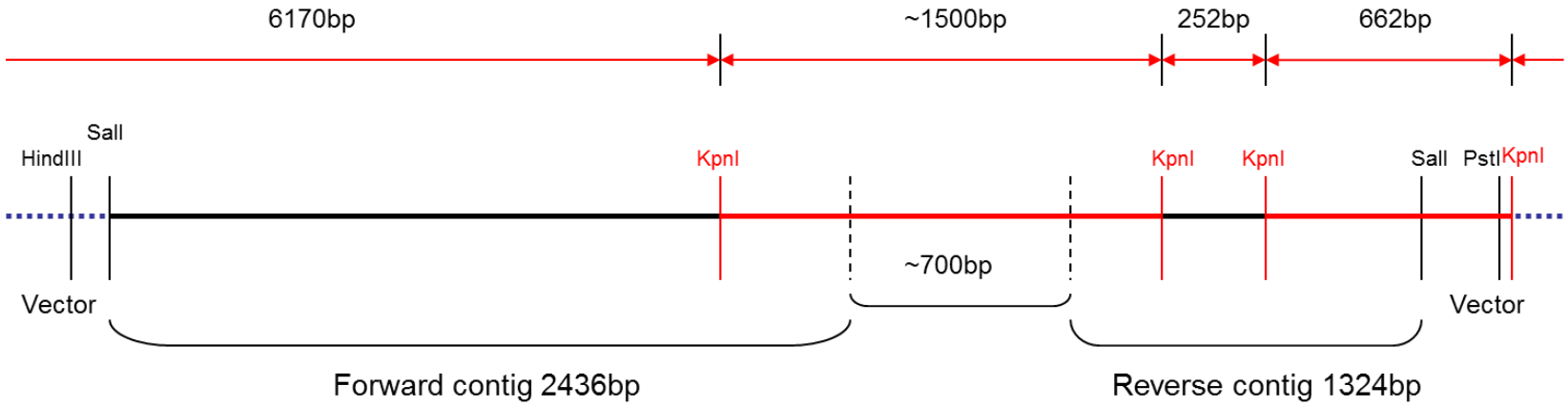


Figure 9: Predicted KpnI restriction map of pS21SatB. Insert is shown in black and vector in blue. The 700 bp gap in the middle is also shown. Seqbuilder produced a map that showed that KpnI would give fragments 6100, ~1500, 662 and ~252bp as shown by the arrow lines on top of the restriction map. Two fragments that were subcloned for sequencing purposes were ~1500bp and 662bp (shown in red).

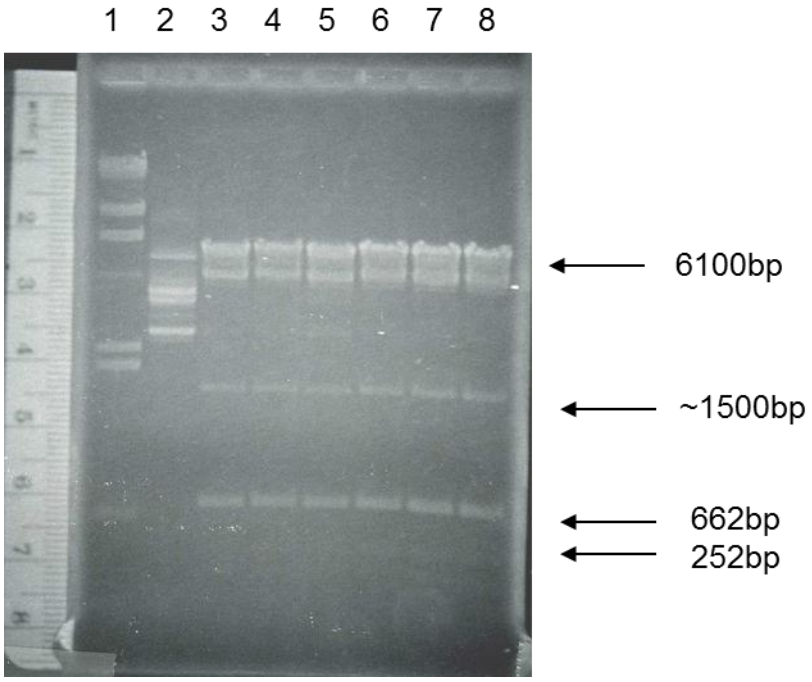


Figure 10: pS21SatB digested with KpnI. Gel showing the restriction digest results of pS21SatB with *KpnI.* Lane 1 is Lambda/*HindIII* marker. Lane 2 pS21SatB (undigested) control. Lane 3-8 pS21SatB digest with *KpnI.* The 252bp band is very faintly visible.

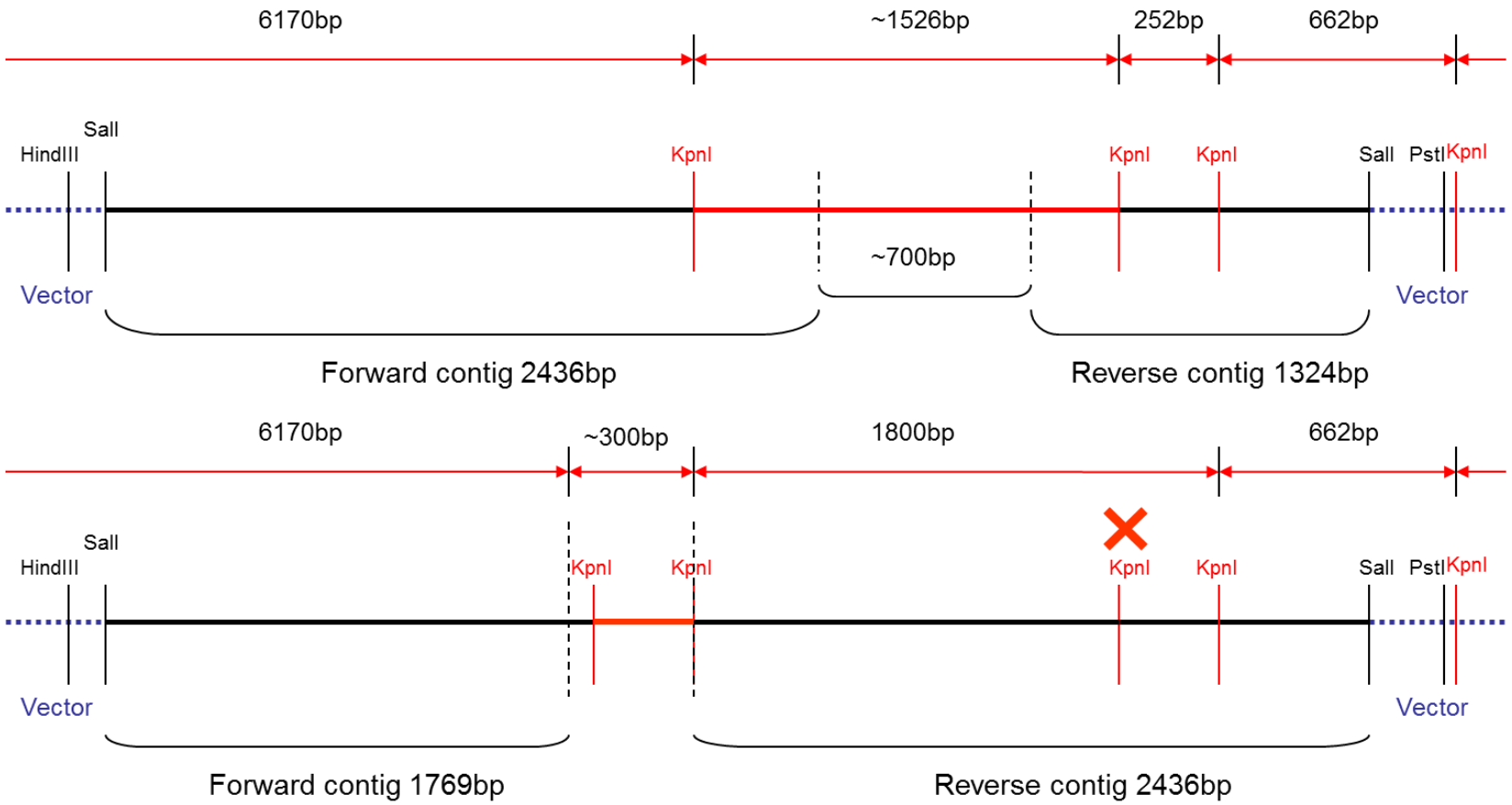
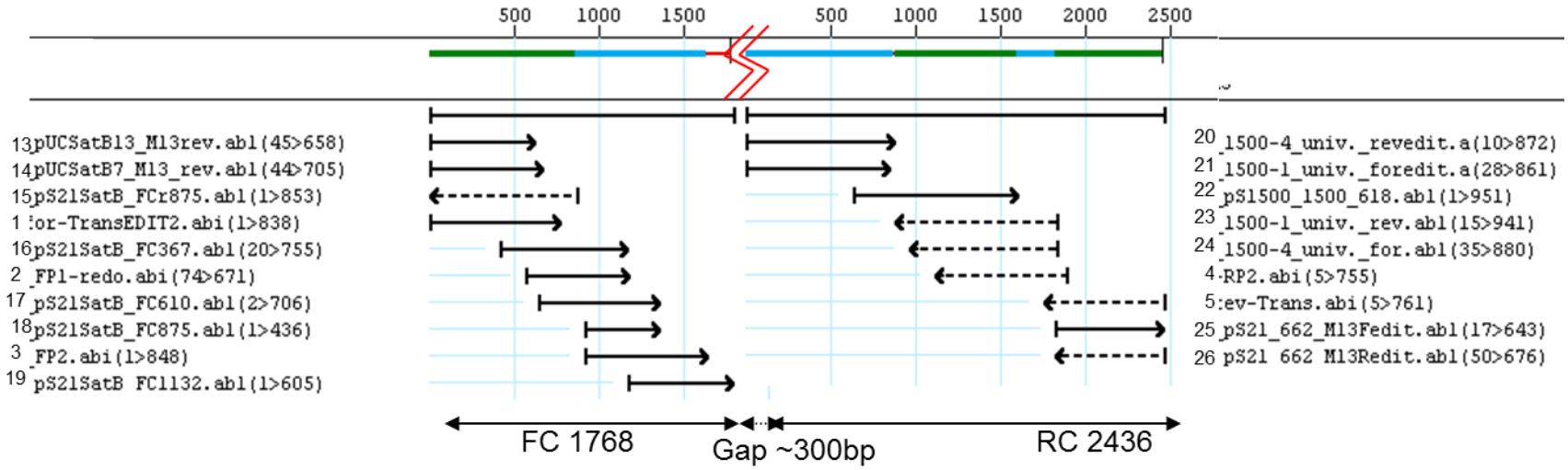


Figure 11: pS21SatB *KpnI* restriction map showing the predicted *KpnI* site and actual *KpnI* site. 11a The predicted 2nd KpnI site (at the start of ~252bp fragment). **11b** 1. The actual size of the ~1500bp fragment was 1800bp. There was false prediction of KpnI site (marked my cross on top) resulting from single poor sequence quality of the reverse contig. The actual KpnI site was then predicted to be in the 300bp gap near the end of the forward contig.



Forward Contig (FC): 1768bp
Reverse Contig(RC): 2436bp
Gap ~300bp
Total size of the clone ~4.8kb

Figure 12: Edited sequence contig of pS21SatB . Forward contig is 1768bp after removing all the transposon generated sequences and reverse contig is 2436bp with a gap of ~300bp.

Sequence #	Position on the contig	Primer Sequence	Sequences from subcloned fragments
1	1	universal primer (forward)	
2	493	TATCAATGTACTCCAACTAT	
3	909	ATCAAAGGACCCAAAATAAAC	
4	4211	TATATTGGATGTGCCTTCA	
5	4778	universal primer (reverse)	
13	1	universal primer (forward)	
14	1	universal primer (forward)	
15	1	TTATTTTGGGTCCTTTGA	
16	419	AAAATATCAAAGTAGCCACAA	
17	639	CTGTACATAAAATAGCAAAGTT	
18	909	ATCAAAGGACCCAAAATAAAC	
19	1165	GTACCCAAACTGTCTATTATTAT	
20	2345	universal primer (reverse)	1500KpnI fragment
21	2345	universal primer (forward)	1500KpnI fragment
22	2975	ACCCAAACTGTGTATTATTA	1500KpnI fragment
23	4146	universal primer (reverse)	1500KpnI fragment
24	4146	universal primer (forward)	1500KpnI fragment
25	4778	universal primer (reverse)	662KpnI fragment
26	4778	universal primer (reverse)	662KpnI fragment

Table 4: List of primers used to build contig shown in Figure 12. The sequence # corresponds to the sequences labeled numerically in Figure 12.

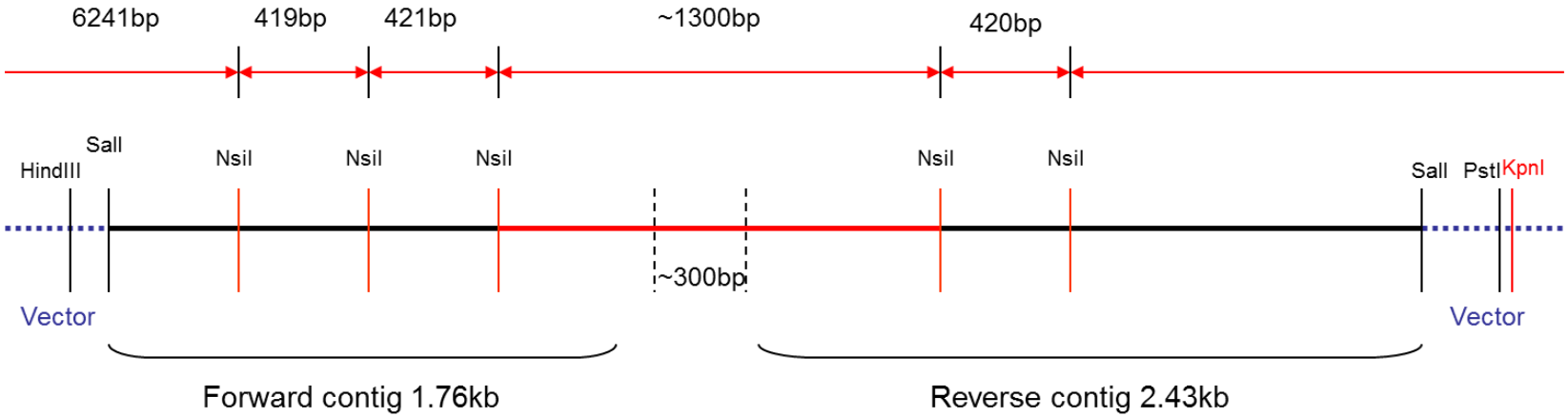


Figure 13: Predicted *NsiII* restriction map of pS21SatB. Schematic showing *NsiI* subcloning fragments: Insert is shown in black and vector in blue. The 300bp gap in the middle is also shown. Seqbuilder produced map showed that *NsiI* would give fragments 6241, 1300 and 3 ~400bp fragments as shown by the arrow lines on top of the restriction map. The fragment that was subcloned for sequencing purpose was ~1300bp (shown in red).

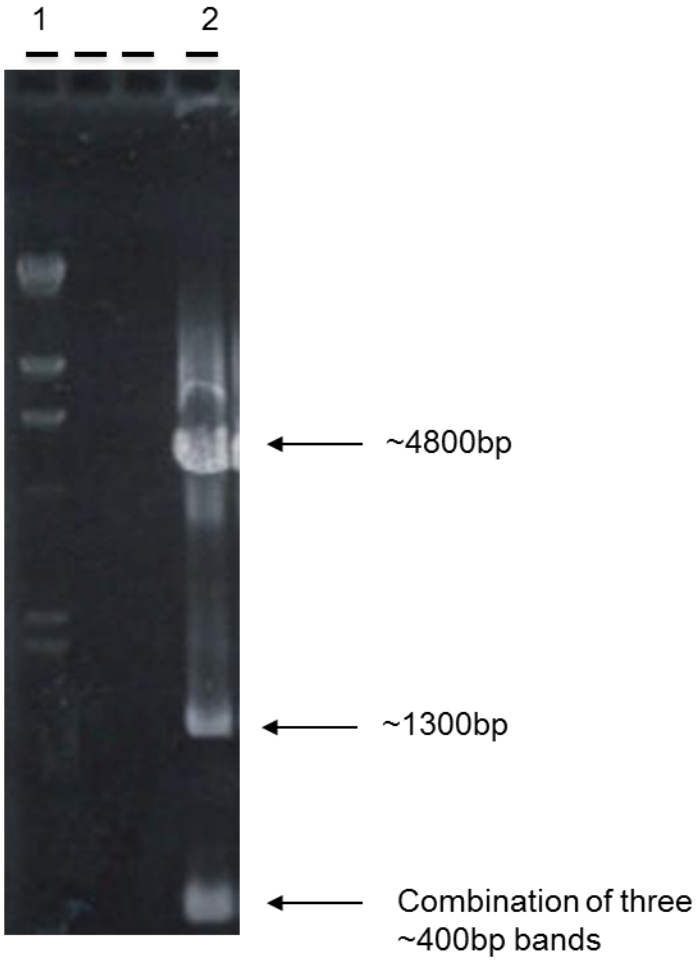


Figure 14: pUCSatB13 digested with NsiI. Gel showing the restriction digest results of pUCSatB13 with *NsiI.* Lane 1 is Lambda/*HindIII* marker. Lane 2 pUCSatB13 digested with *NsiI.* The ~1300bp band used for sub-cloning is visible. Since pUC was the vector used for cloning, the first band in lane 2 is 4.8kb and not 6.2kb as shown in Figure 7.

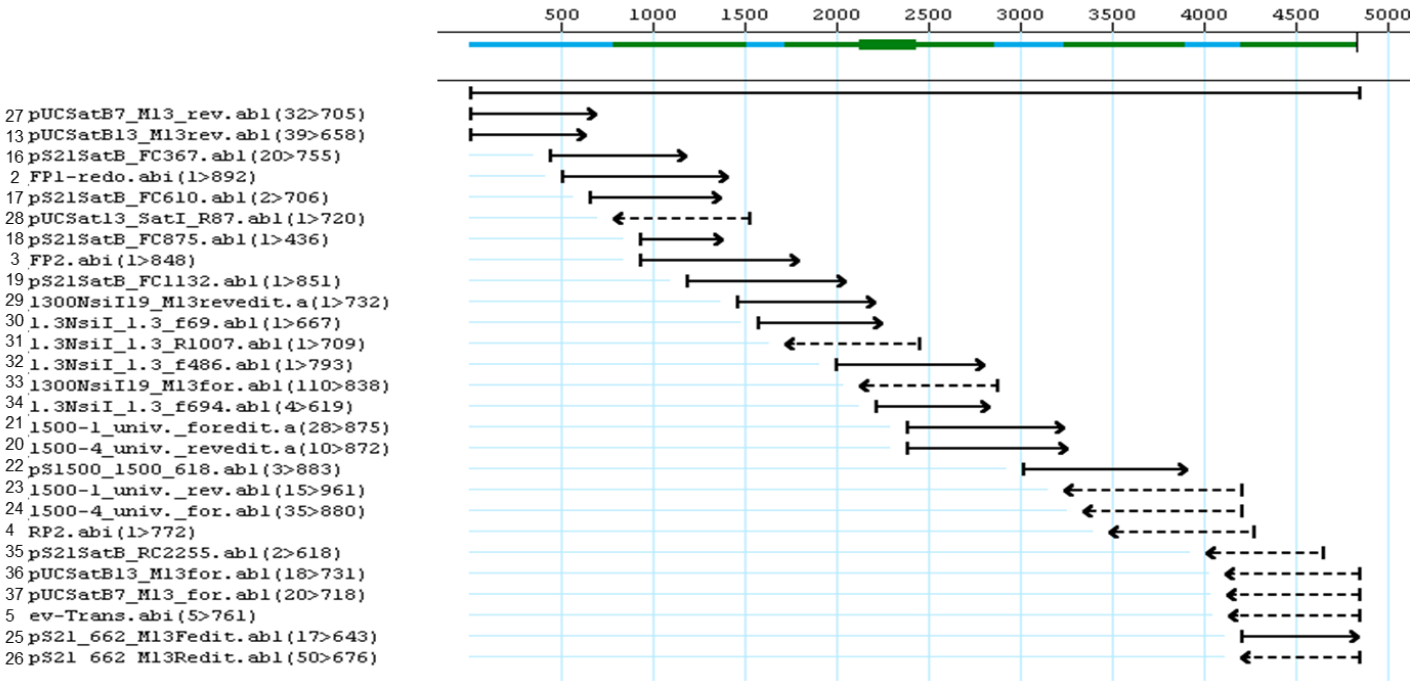


Figure 15 : Completed sequence contig of pS21SatB. pS21SatB is 4778bp in size and the clone has double coverage over its entire length.

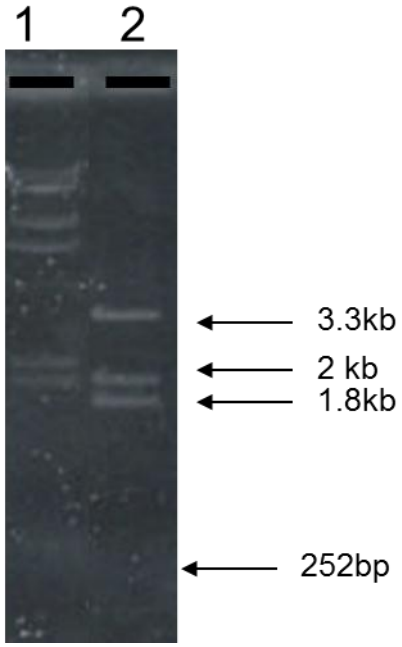


Figure 16: pUCSatB13 digested with KpnI. Gel showing the restriction digest results of pUCSatB13 with *KpnI.* Lane 1 is Lambda/*HindIII* marker. Lane 2 pUCSatB13 digested with *KpnI.* The 3.3kb(2686bp vector + 662bp insert), 2kb (insert fragment),1.8kb (insert fragment) fragments are clearly visible. The 252bp(insert fragment) band is faintly visible. Note that the sizes of bands are different from Figure 4 because pUC is used as a vector in cloning pUCSatB13 instead of pSPORT which was used for cloning pS21SatB.

Sequence #	Position on the contig	Primer Sequence	Sequences from subcloned fragments
2	493	TATCAATGTACTCCAACTAT	
3	909	ATCAAAGGACCCAAAATAAAC	
4	4211	TATATTGGATGTGCCTTCA	
5	4778	universal primer (reverse)	
13	1	universal primer (forward)	
16	419	AAAATATCAAAGTAGCCACAA	
17	639	CTGTACATAAAATAGCAAAGTT	
18	909	ATCAAAGGACCCAAAATAAAC	
19	1165	GTACCCAAACTGTCTATTATTAT	
20	2345	universal primer (reverse)	1500KpnI fragment
21	2345	universal primer (forward)	1500KpnI fragment
22	2975	ACCCAAACTGTGTATTATTA	1500KpnI fragment
23	4146	universal primer (reverse)	1500KpnI fragment
24	4146	universal primer (forward)	1500KpnI fragment
25	4778	universal primer (reverse)	662KpnI fragment
26	4778	universal primer (reverse)	662KpnI fragment
27	1	universal primer (forward)	
28	1487	TATATTTGATGAACCTACA	
29	1437	universal primer (reverse)	1300NsiI fragment
30	1544	TGTAGGTTCATCAAATATA	1300NsiI fragment
31	2409	TATATTGGATGTGCCTTCA	1300NsiI fragment
32	1965	TTAAAGTTGCCCATATATACATTCTA	1300NsiI fragment
33	2824	universal primer (forward)	1300NsiI fragment
34	2176	TATCAAAGTACACCAAGT	1300NsiI fragment
35	4583	TATTCGGGTACTTTGATTTA	
36	4778	universal primer (forward)	
37	4778	universal primer (forward)	

Table 5: List of primers used to build contig shown in Figure 15. The sequence # corresponds to the sequences labeled numerically in Figure 15.

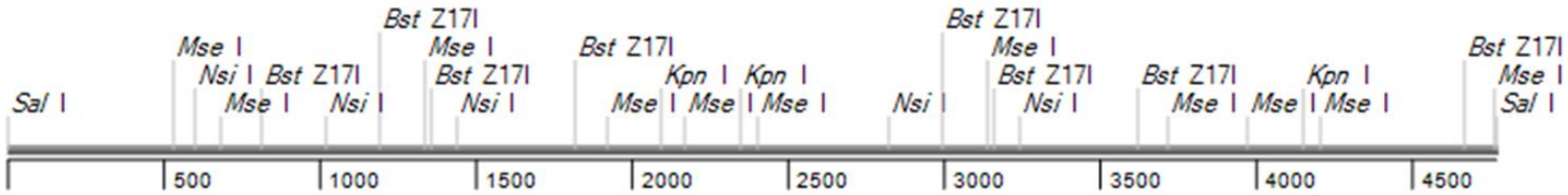


Figure 17: Restriction map of pS21SatB. All the restriction enzymes used for sequencing and studying pS21SatB are shown above.

GTCGACCAAATATATATTATATACTGTGCATAAAATATGAAGGTACATCAAATATATACTATATGCTGCACATAAAATATCAAAGTACCCAAAGTATGTATTATATACT
GTACATAAAATATCAAAGTACCCAAAGTATATATTATATACTGTACATAAAATATAAAATACCCCAAATATATATTATATGCTGTACATAAAATATCAAAGTACACAAAA
TATTTATTATATACTGTACGTAAAATACCAAAGGACCCAAAATATATATTATATACTGTAGATAAAATATCAAAATTCCCAAAATATATATTATATACTGTACATAAAATA
CCAAAGTACCCAAAGTATATATTATATACCGTACAAAAAATATCAAAGTAGCCACAATATGTATTATACACTGTACATAATATATGAAAGTACCCAAACATTTATAATAA
ACTGTACATAAAATATCAATGTACTCCAACTATATATTATGTGCGGTACATAAGATATCAAAGTACCCATACTATATAGTATATACTGTACTTAAAATATCAAAATACCC
AAAATATATATTATATACGTTACATAAACTATCAAATACACAAAGTATGCATTATATACTGTACATAAAATAGCAAAGTTCCCAAAATATGTATTATATACTGTACATAA
AATATGAAAGTACCCAACCATATTTAATAAACTCTACATAAAATACCAAAGTACCCAAACTGTGTATTATATACTGTCCATAAAATATCAAAGTACCCAAAATATATATT
GTAAACTCTACATAAAATATCAAAGTAGCCAAAATACATATTGTATACTGTACATAAAATATAAAAGTACCCCAAATATATATTATATACTGTACATAAAATATCAAAGG
ACCCAAAATAAACATTATATACTGTACATAAAATATCCAAATACTCAAAATATATATTACATACTGTACATAAAATATCAAAGTACCCAAATTATGTATTATATACTGTAC
GTAAAATATCAGAGTACCCAAAATATGCATTATATACTGTACATAAAATATGAAAGTAACAAAACATTTATAATAAACTGTACATAAAATATCAAAGTACTCAAACTATA
TATCATATACTGTACATAAAATATCAAAGTACCCAAACTGTCTATTATTATATACTGTATATAAAATATCAAAGTACCTGAAATATATACTGTATACTGTACATAAAATAT
CAAAGTACCCAAAATATATACTATATACTGTACATAAAATATGAAGGTACATCAAATATATATTATATCCTGTACATAAAATATCAAAGTACACCTATTATATATTGTGT
ACTGTACATAATATTTTAAATCGCCCAAATATATATTGTATACTGTATATAAAATATGAAGGTACTTCAAATATATATTATATGCTGCACATAAAATATCAAAGTACCCA
AAGTATGCATTACATACTGTACATAAAATATCAAAGTACACAAATTATATATTATATACTGTACATAAAATATGTAGGTTCATCAAATATATTTTATATTCTGTACATAAT
ATATCAAAGTACATCAAATATATATTATATACGGTACATAAAATATCAAAGTACCCAAAATATGTATTATATACTGTACATAAAACATGAAAGTAACCAAACATTTATAA
TGAACTGTACATAAAATATCAAAGTACTCAATCTACATATTATATACACTATATAAAATATCAAAGTACACAAACTGTATATTATATACTGTACATAATATATCAAAGTA
CCCAGAATATATAGCATACTGTACATAAAATATCAAAGTACCCAAAATATATATTGTATACTGTACATAAAATATGAAGGTACATCAAAGATATATTATATTCTGTACAT
AAAATATCAAAGTACACCAAATATATATTATATGCTGTACATAATATATTAAAGTTGCCCATATATACATTCTATACTGTACATAAAATATGAAGGTACATCAAATATAT
AGTATATGCTGCACATAAAATATCAAAGTACCCAAAGTATGTATTATACACTGTACATAAAATATCAAAGTACCAAAAGTATATATTACATACTGTAAATAAAATATGAA
GGTACCTCAAATATATATTATGTTCTGTACATAAAATATCAAAGTACACCAAGTATATACTATATACTCTGCATAAAATATTAAATTACCCAAAATATGTATTGTATTGT
GTACATAAAATATGAAAGTAACCAAACATTTATAATAAACTGTACATAAGATATCAAAGTACTCAAACTGTATATTATATACTGTACATAAAGTATCAAAGTACCCAAAC
TGTGTAGTATATACTGTACATAAAATATCAAGGTACCCAAAATATATATTGTAAACTGTAGATAAAATATCAAAGTACCCAAAATATATATTAAATACTGTACATAAAAT
ATGAAGGCACATCCAATATATATTATATTCTGTACATAAAATATCAAACTACACCAAATGTATATTATATACTGTACATAAAATATCAAAGTACCCAAATGTATATTATA
TACTGTACATAAAATATCAAAGTCCCCAAAGTGTATATTATATACTGTACATAAAATATGAAGGTTCATCAAATATATTTTATATTCTGCACATAATATATCAAAGTACA
GCAAATATATATTATATACTGCACATAAAATATCAAAGTACCCAAAATATGTATTATATACTGTACATAAAATATCCAAATACTCAAAATATATATTATATACTGTACATA
AAATATCAAAGTACCCAAATTATGTATTATATACTGTACGTAAAATATCAGAGTACCCAAAATATGCATTATATACTGTACATAAAATATGAAAGTAACAAAACATTTAT
AATACACTGTACATAAAATATCAAAGTACTCAAACTATATATCATATACTGTACATAAAATATCAAAGTACCCAAACTGTGTATTATTACATACTGTATATAAAATATCA
AAGTACCTGAAATATATATTGTATACTGTACATAAAATATCAAAGTACCCAAAATATATACTATATACTGTACATAAAATATGAAGGTACATCCAATATATATTATATCC
TGTACATGAAATATCAAAGTACACCTATTATATATTGTGTACTGTACATAATATATTAAATCGCCCAAATATATATTGTATACTGTATATAAAATATGAAGGTACTTCAA
ATTCATATTACATGCTGCACATAAAAATCAAAGTACCCAAAGTATGCATTACATACTGTACATAAAATATCAAAGTACACAAATTATATATTATATACTGTACATAAAAT
ATGAAGGTTCATCAAATATATTTTATATTCTGTACATAATATATCAAAGTACATCAAATATATATTATATACTGTACATAAAATATCAAAGTACCAAAAATATGTGTTATA
TACTGTATATAAAATATGAAAGTAACCAAACATTTATAATAAATTGTACATAAAATATGAAAGTACTCAATCTACATATTATATACAGTATATAAAATATCAAAGTACCC
AAACTGTATATTATATACTGTACATAATATATCAAAGTACCCAGAATATATAGCATACTGTACATACAATATCAAAGTACCCAGAATGTATATTGTATACTGTACATAAA
ATATGAAGGAACATCAAAGATATATTATATTCTTTACATAAAATATCAAAGTACACCAAATATATATTATATGCTGTACATAATATATTAAACTTGCCCATGTATAGATTT
TATACTGTACATAAAATATGAAGGTACATCAAATATATAGTATATGCTGCACATAAAATATCAAAGTACCCAAAGTATGTATTATATACTGTACATAAAATATCAAAGTA
CCAAAATTATATATTACCTACTGTAAATAAAATATGAAGATACCTCAAATATATATTATATTCTGTACATAAAATGTCAAAGTACACCAAGTATATACTATATACTCTACA
TAAAATATTAAATTACCCAAAATATGTATTGTATTGTGTACATAAAATATGAAAGTAACCAAACATTTATAATAAACTGTACATAAAATGTCAAAGTACTCAAACTGTAT
ATTATATACTGTACATAAAGTATCAAAGTACCCAAACTGTGTAGTATATACTGTACATAAAATATCAAGGTACCCAAAATATATATTGTAAACTGTAGATAAAATATCAA
AGTACCCAAAATATATATTAAATACTGTACATAAAATATGAAGGCACATCCAATATATATTATATTCTGTACATAAAATATCAAAGTACACCAAATGTATATTATATACT
GTACATAAAATATCAAAGTCCCCAAAGTGTATATTATATACTGTACATAAAATATGAAGGTTCATCAAATATATTTTATATTCTGCACATAATATATCAAAGTACAGCAA
ATATATATTATATACTGTACATAAAATATCAAAGTACCCAAAATATGTATTATATACTGTACATAAAATGGGAAAGTAACCAAACATTTATAATAAACTGTACATAAAAT
ATCAAAGTACTCAAACTACATATTATATACAGTATATAAAATGTCAAAGTACCCAAACTGTATATTATATACTGTACATAATAAATCAAAGTACCCGAATATATAGCATA
CTGTACATAAAATATCAAAGTACCCAAAATATATATTGTATACTGTACATAAAATATGAAGGTACATCAAAGATATATTATATTCTGTACATAAAATATCAAAGTACACC
AAATATATATTATATGCTGTACAGAATATATTAAAGTCGAC

Figure 18: pS21SatB sequence. The completed 4778bp of the pS21SatB sequence are shown beginning and ending at *SalI* restriction site.

	pTR1-6	pN6.4	Y-specific Sat I On HC21	HC4 sat I
pS21SatB	1. 95% 7 gaps 2248/2346 2. 83% 32 gaps 2494/2977 3. 89% 16 gaps 2678/2981	82% 22gaps 499/605	----	83% 35gaps 881/1059

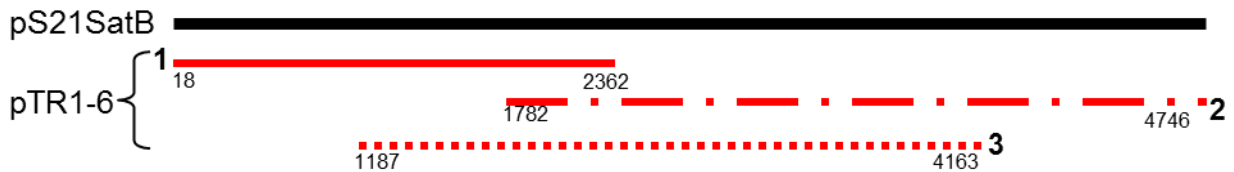


Table 6: BLAST2seq results of pS21SatB against all known satI sequences. The %matches are the highest match over the longest distance. The fractions indicate the total number of base pairs that are similar in both the sequences over the total length of the region compared.

Accession # and sequence description	% highest match to pS21SatB over the longest distance	% highest match to pS21SatB
L01057.1 Human (clone TRI-6) satellite I repeat region	2714/3011 (90%)	607/629 (96%)
AC137488.2 Homo sapiens chromosome 16 clone CTD-2506I5, complete sequence	3672/4590 (80%)	407/499 (81%)
AC022616.9 Homo sapiens chromosome 8, clone RP11-726G23, complete sequence	2895/3552 (81%)	49/54 (90%)
U10629.1 Human clone pSVH1-28 satellite sequence from chromosome 4 centromere	905/1079 (83%)	146/164 (89%)
U05290.1 Human middle repetitive DNA sequence L2HS6439	508/615 (82%)	499/603 (82%)

Table 7: Whole genome BLAST results of pS21SatB. The top 5 sequence matches are indicated.

Parameters: 2 7 7 80 10 50 2000
Organism:
sequence: pS21SatB (more info)
FASTA header:

Genbank:
Sequence length: 4778
Date processed: 2008-09-08

Indices	Period Size	Copy Number	%Matches	%Mismatches	%Indels	%A	%C	%G	%T
6--4774	41	114.000000	88	9	3	46	14	8	31

Table 8: Tandem Repeats Finder results for pS21SatB. This table shows that the pS21SatB satellite I monomer is 41bp in size and there are 114 monomers in pS21SatB. %match to the consensus is 88% as shown in the table above.

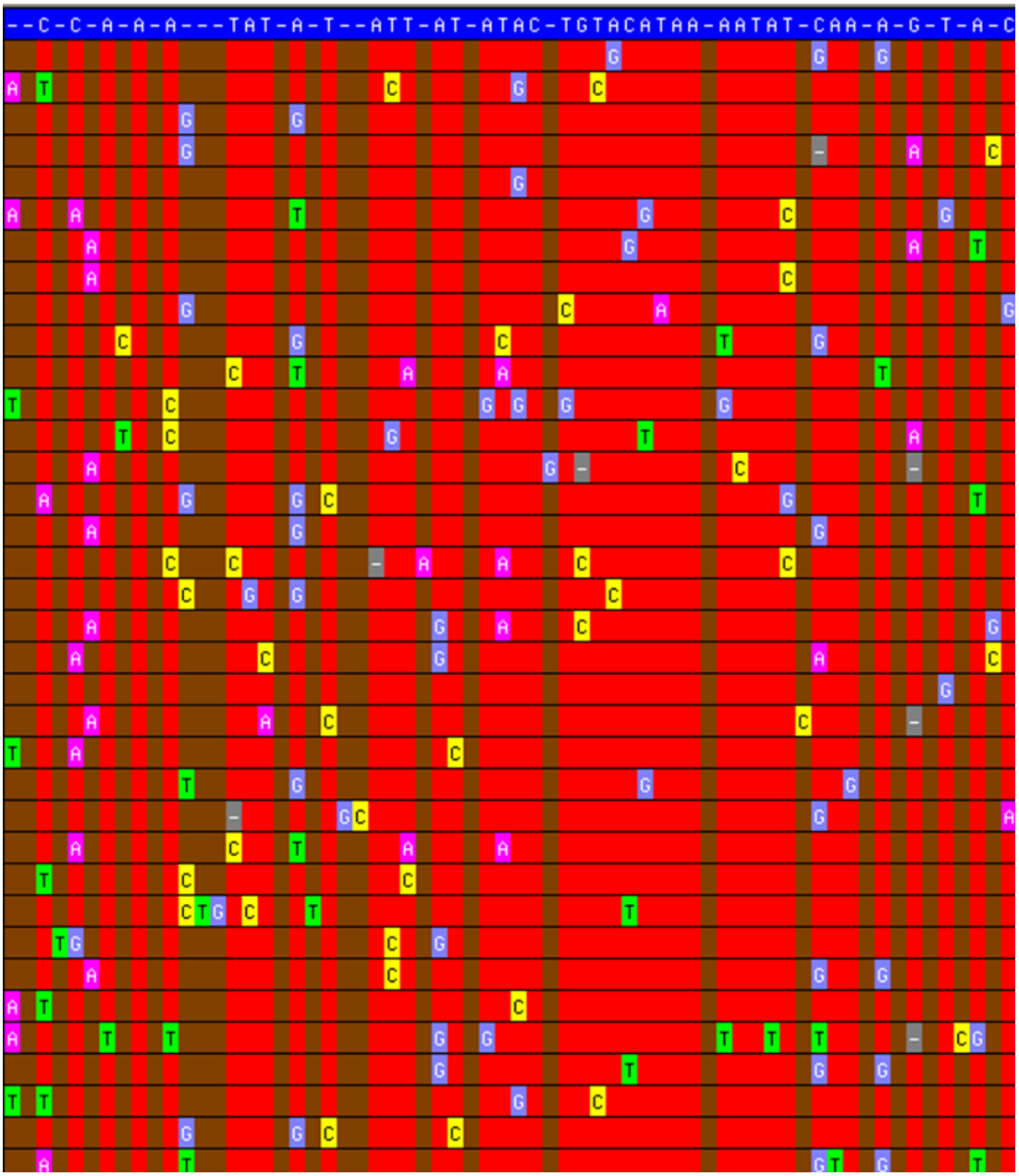
Comparison of pS21SatB consensus vs. pTR1-6 consensus:

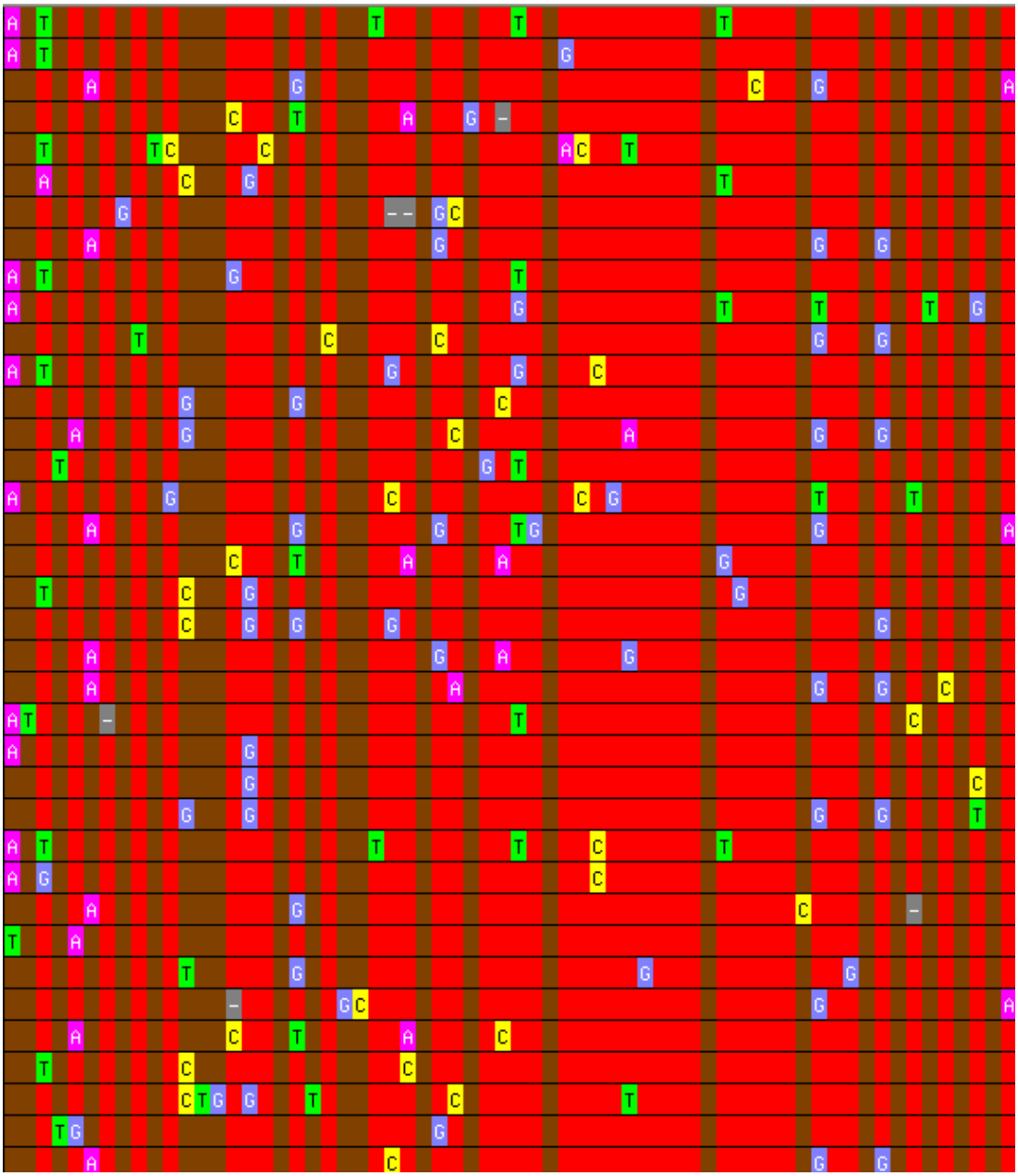
1. pS21SatB consensus (41bp):
   CCAAATATATATTATATACTGTACATAAAATATCAAAGTAC

2. pTR1-6 sequence as reported in Kalitsis et al 1992(42bp):
   GTACTTTGATATTTTATGTACAGTATATAATATATATTTTGG
   Reverse and complement of this sequence:
   CCAAAATATATATTATATACTGTACATAAAATATCAAAGTAC

3. Comparison of pS21SatB sequence against reverse and complimented pTR1-6 sequence:
   CCAAA__TATATATTATATACTGTACATAAAATATCAAAGTAC
   CCAAAATATATATTATATACTGTACATAAAATATCAAAGTAC

Figure 20: Comparison of pS21SatB vs. pTR1-6 consensus sequences. GTAC is the *RsaI* restriction site which was used by Kalitsis et al (1992) to line up the monomers. Here it serves as a reference point to line up the 2 different consensus sequences. 1. shows the consensus sequence of pS21SatB obtained from tandem repeat finder. 2. shows the reported pTR1-6 sequence and its reverse and compliment. 3. shows the comparison of pS21SatB vs. pTR1-6 consensus sequence. Except for the 6th nucleotide "A" in 3(shown as "_" in pS21SatB) both consensus sequences are the same.


--C-C-A-A-A---TAT-A-T--ATT-AT-ATAC-TGTACATAA-AATAT-CAA-A-G-T-A-C



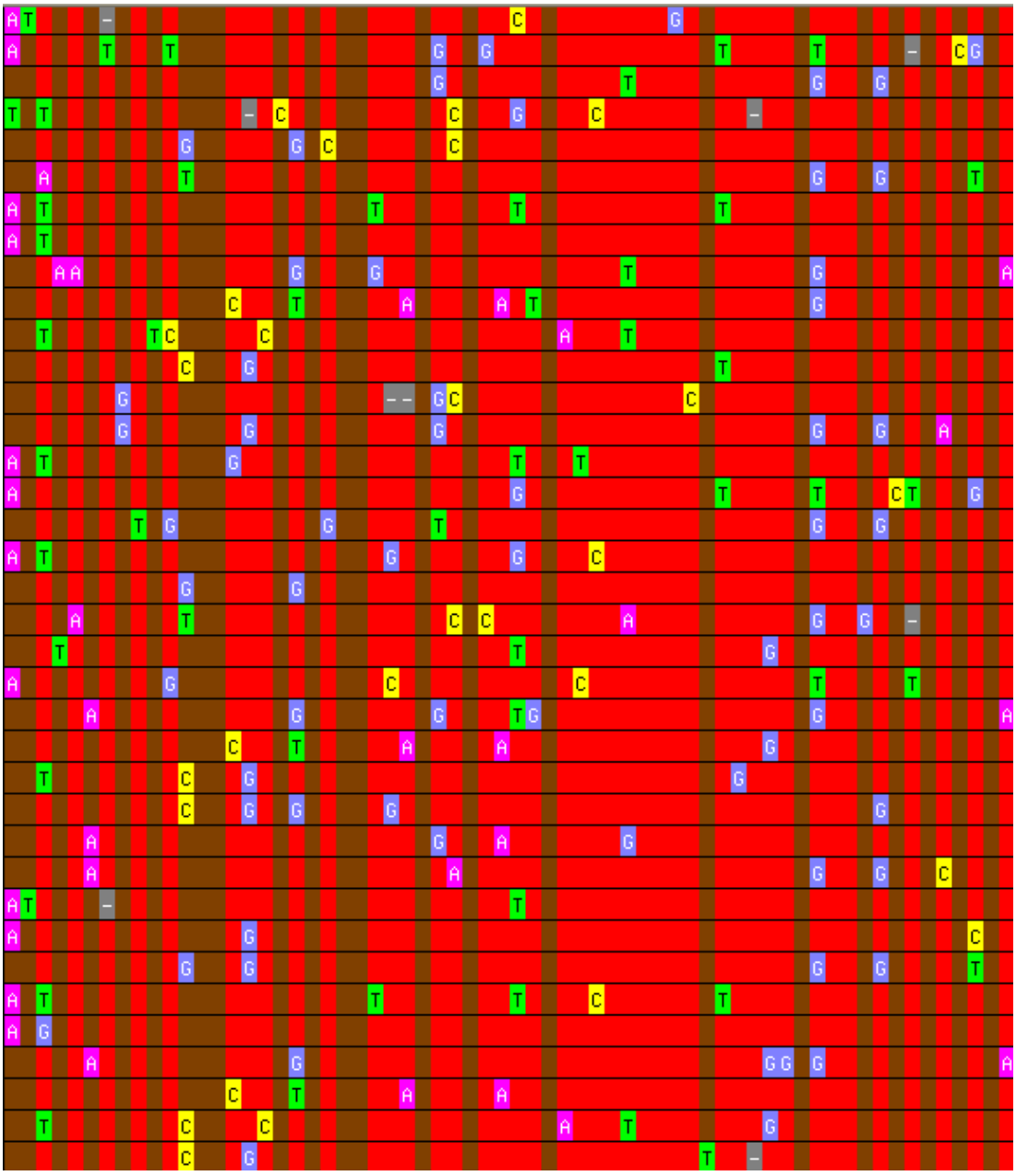


Figure 21: Tandem Repeat Finder results. All 114 monomers are lined up with the consensus sequence shown at the beginning.

	1	2	3	4	5	6	7	8	9	10	11
1	100%	83%	83%	87%	85%	84%	85%	83%	81%	83%	83%
2		100%	85%	80%	80%	80%	85%	79%	79%	80%	80%
3			100%	81%	81%	82%	83%	80%	80%	81%	82%
4				100%	83%	80%	82%	77%	82%	80%	81%
5					100%	82%	83%	79%	79%	83%	82%
6						100%	83%	81%	80%	82%	81%
7							100%	82%	79%	82%	84%
8								100%	79%	80%	79%
9									100%	80%	79%
10										100%	89%
11											100%

Table 9: Internal BLAST analysis results of pS21SatB. The sequence was divided into eleven segments of 435bp each. BLAST analysis determined the highest %similarity over the longest distance for each segment compared to all others.

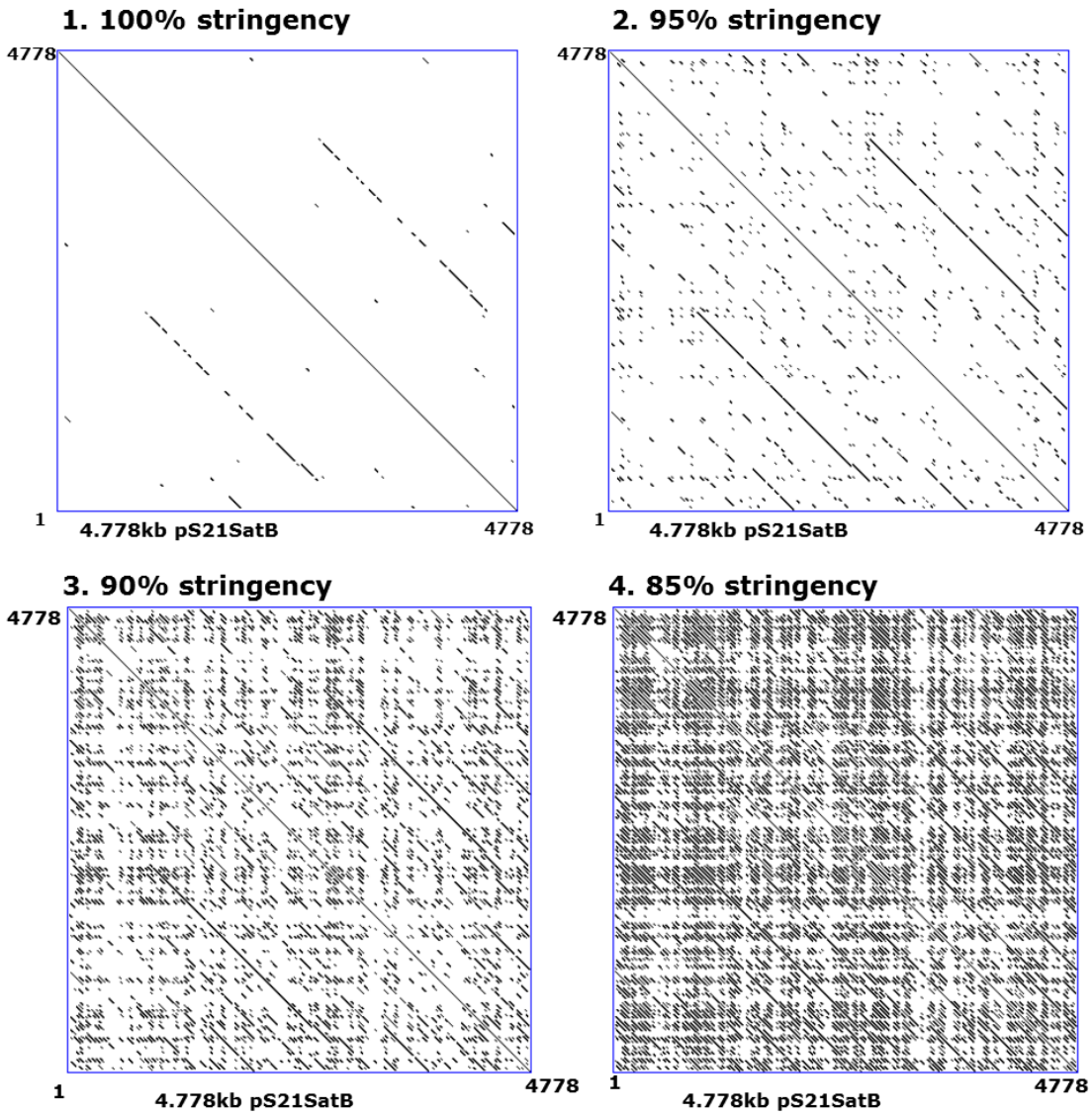


Figure 22: Dotplots of pS21SatB at different stringencies. The parameters used for 1. 100% stringency-window size 41, mismatch limit 0. 2. 95% stringency-window size 41, mismatch limit 3. 3. 90% stringency-window size 41, mismatch limit 4. 4. 85% stringency- window size 41, mismatch limit 6.

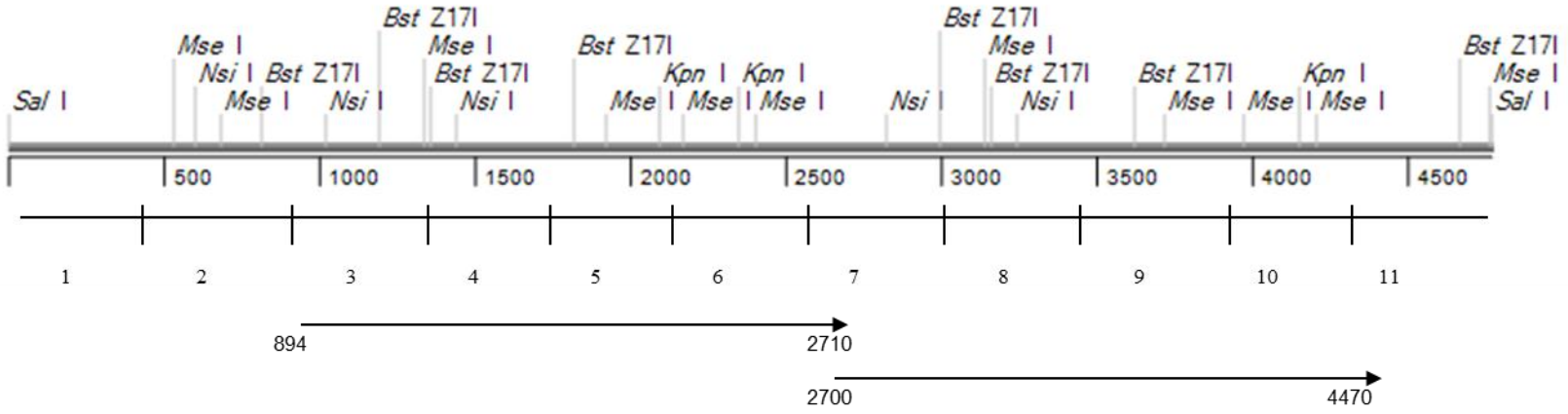


Figure 23: pS21SatB direct repeat location. The restriction map and 11 segments of pS21SatB are shown. The direct repeats are indicated by arrows. The first repeat begins at nucleotide position 894 and ends at 2710. The second repeat begins at nucleotide position 2700 and ends at 4470. There is a 10bp overlap between the repeats.

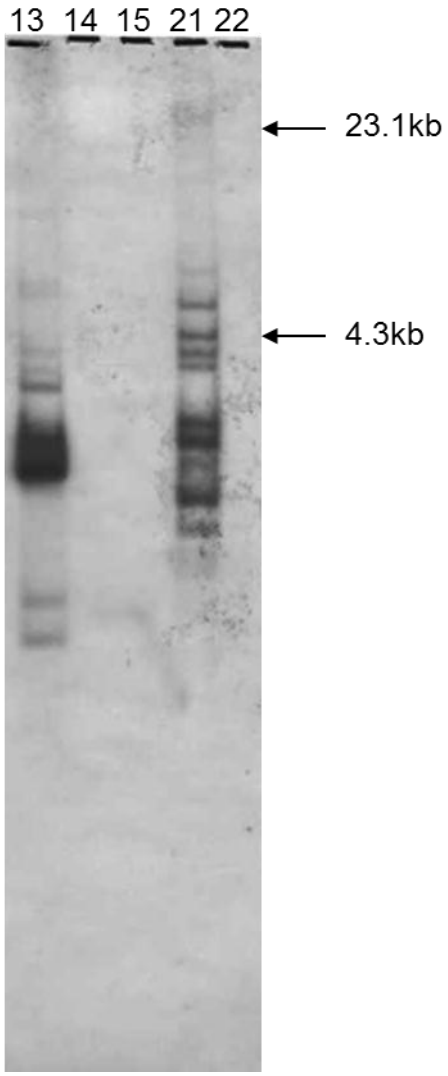


Figure 24: Autoradiogram of acrocentric blot with pS21SatB. DNAs from hybrid cell lines containing each of the acrocentric chromosomes (13,14,15,21 and 22) separately were cut with KpnI. Hybridization was done at 97% stringency(0.15XSSC at 62.4OC). There is no 3.0 kb higher order repeating band on HC21, as seen for HC13, indicating that pS21SatB does not have a higher order repeating structure.

Fragment #	Nucleotide position on pS21SatB	Nucleotide position on pTR1-6	%match to pTR1-6
1	1-435	1912-2340	96
2	436-870	1477-1911	98
3	871-1305	1005-1434	91
4	1306-1740	610-1041	96
5	1741-2175	176-609	97
6	2176-2610	1305-1716	86
7	2611-3045	1062-1494	90
8	3046-3480	669-1103	99
9	3481-3915	237-668	97
10	3916-4350	2310-2740	87
11	4351-4778	2344-2766	96

Table 10: BLAST analysis results of 11:435bp pS21SatB segments to pTR1-6. The maximum % similarity over the longest distance of each pS21SatB fragment against pTR1-6 is reported.

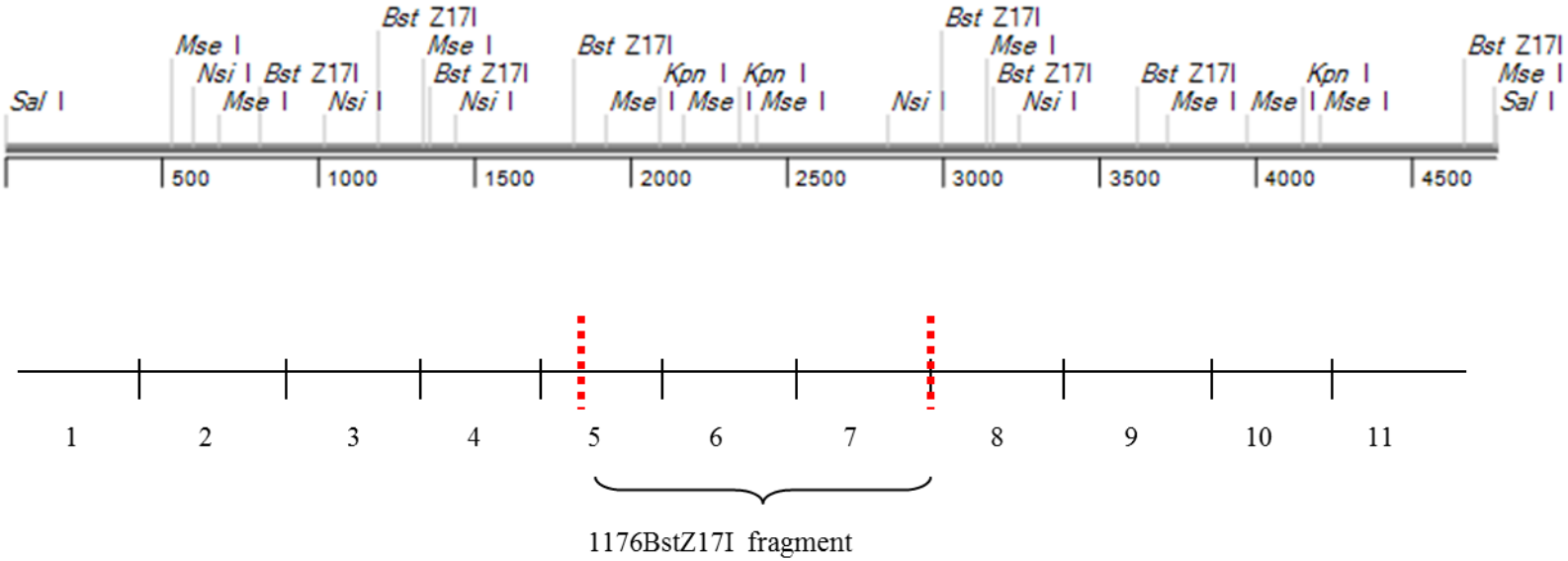


Figure 25: Location of 1176bp BstZ17I fragment. Schematic showing pS21SatB restriction map. Below that is a schematic showing the 11 fragments of 435bp each and the putative region from where an HC21 specific sequence could be obtained. The 1176bp *BstZ17I* fragment is shown in red.

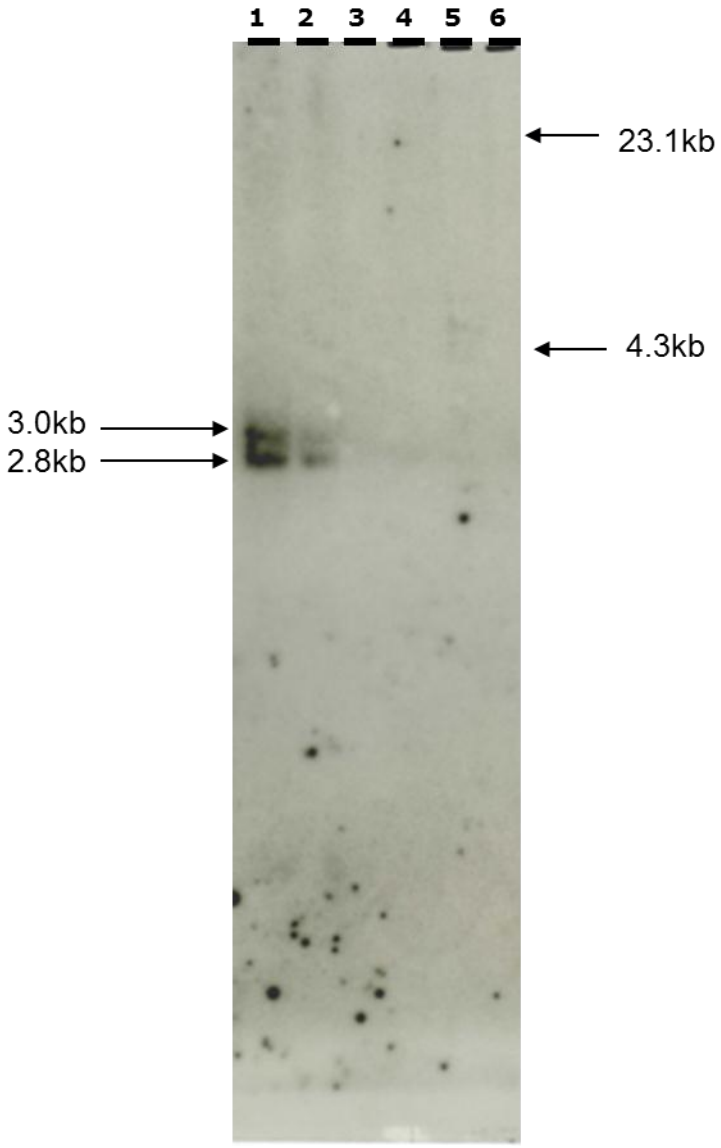


Figure 26: Autoradiogram of acrocentric blot with 1176BstZ17I. DNAs from various acrocentric chromosomes were digested with KpnI and run on a 1% agarose gel. The gel was then blotted and hybridized with 1176BstZ17I at 95% (0.2XSSC, 62.4^{O}C) stringency. The lanes are 1. control (total human DNA), 2. HC13, 3. HC14, 4. HC15, 5. HC21 and 6. HC22.

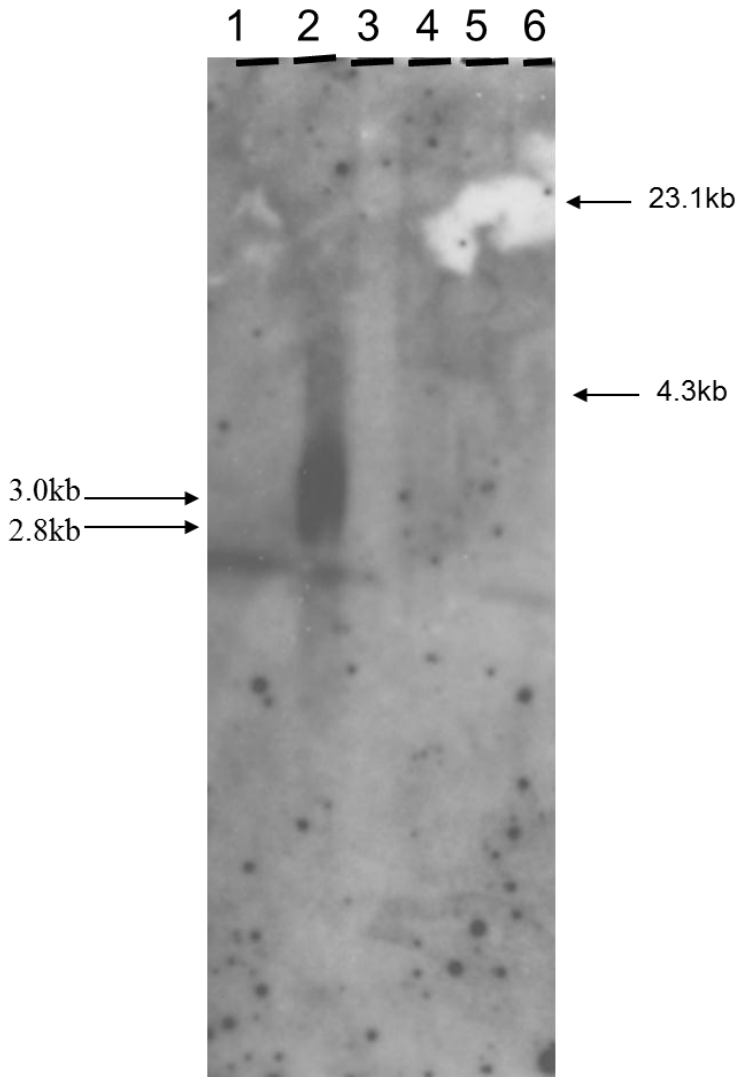


Figure 27: Autoradiogram of acrocentric blot with 1176BstZ17I. DNAs from various acrocentric chromosomes (Coriell Repository) were digested with *KpnI* and run on a 1% agarose gel. The gel was then blotted and hybridized with 1176BstZ17I at 95% stringency. The lanes are 1. control (total human DNA), 2. HC13, 3. HC14, 4. HC15, 5. HC21 and 6. HC22.

Figure 28: BLAST2seq alignment of 1176bp BstZ17I against pTR1-6. The start and end of the highly similar region is indicated by arrows.

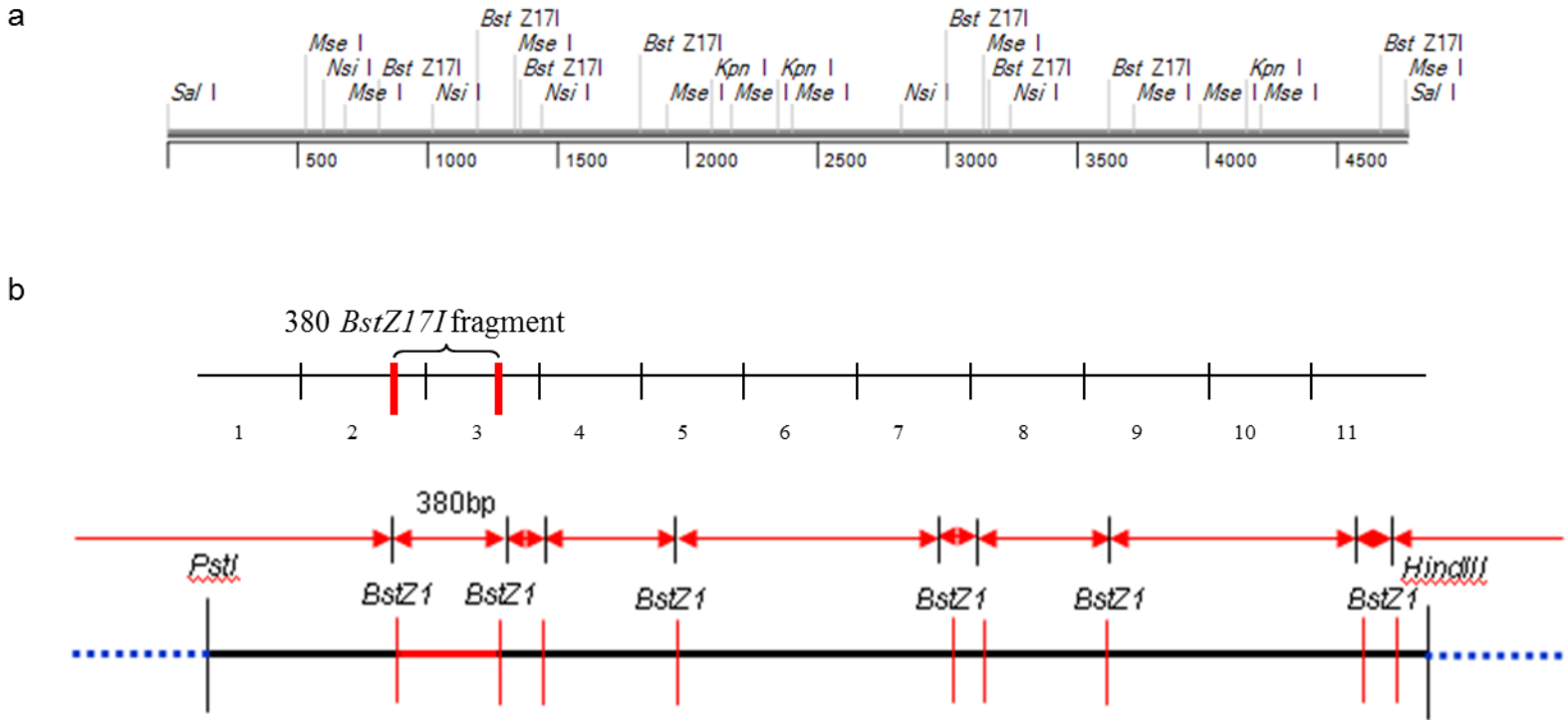


Figure 29: Putative HC21 specific sequence from pS21SatB. a. A restriction map of pS21SatB **b.** Location of 380bpBstZ17I fragment relative to the 11 fragments is also shown. This probe overlaps segment 2 a little and covers most of segment 3. Restriction map of pS21Satb when digested with BstZ17I is shown. The insert is shown in black, the vector is in blue and the 380bp fragment is shown in red.

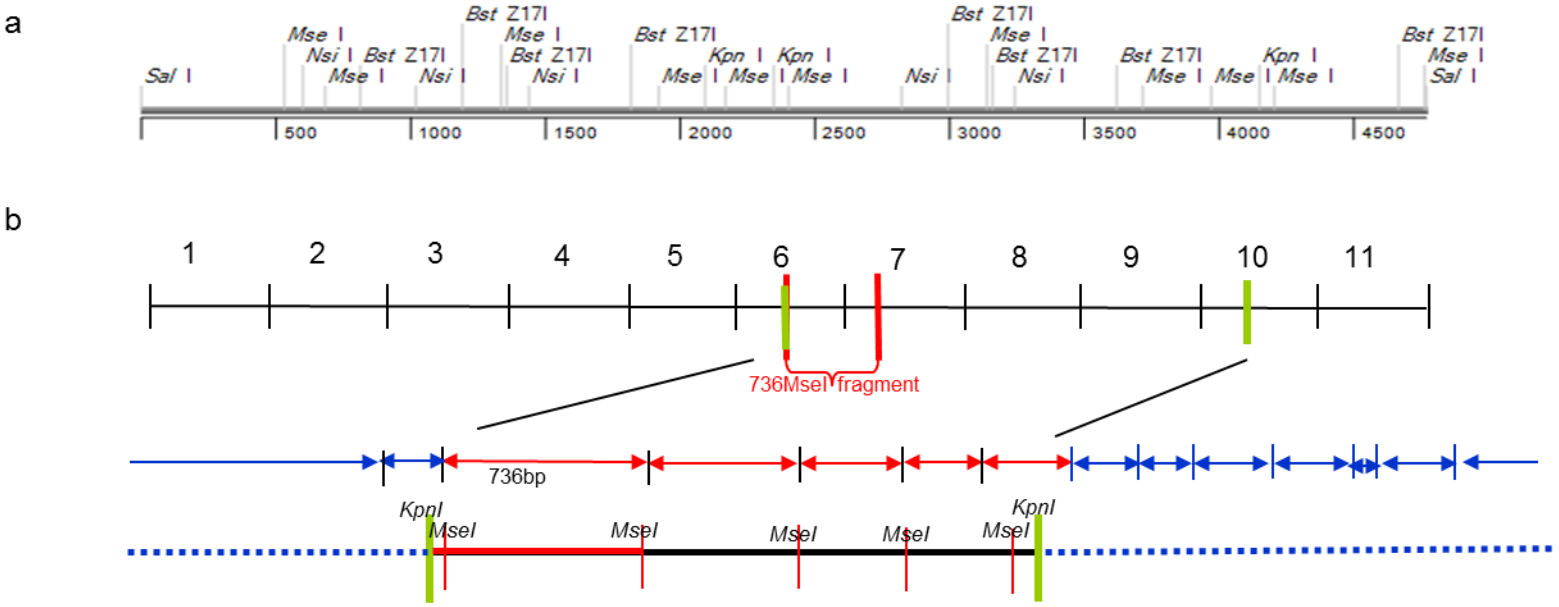


Figure 30: Putative HC21 specific sequence from pS21SatB. a. A restriction map of pS21SatB. **b.** Location of 736bp*MseI* fragment *(*1. & 2.)corresponding to the 11 fragments. Restriction map of 1800bp*KpnI* fragment digested with *MseI* is shown (3.). *736bpMseI* fragment was gel extracted from sub-cloned 1800bp*KpnI* fragment. The insert is shown in black, the vector is in blue and 736bp fragment is shown in red.

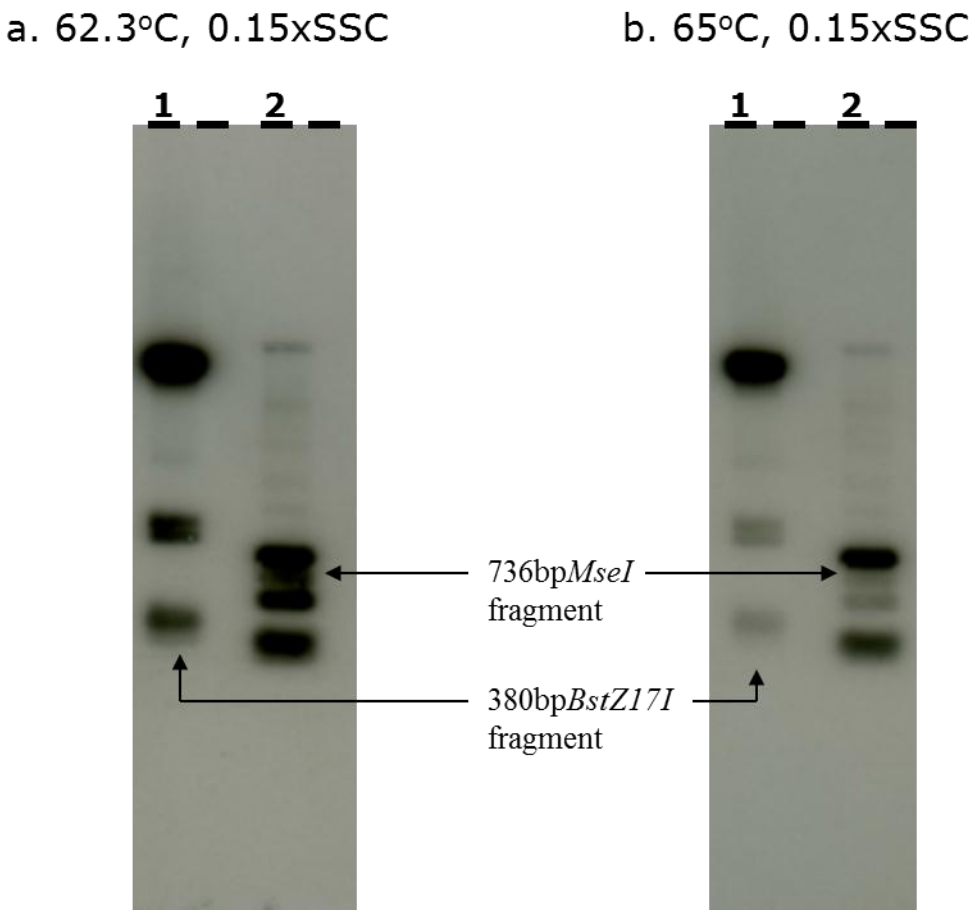


Figure 31: Autoradiograms of minigel blot. Lane 1: pUCSatB digested with *BstZ17I* and Lane 2: 1800*KpnI* fragment digested with *MseI.* The blot was hybridized to pTR1-6 at the stringencies indicated for each autoradiogram.

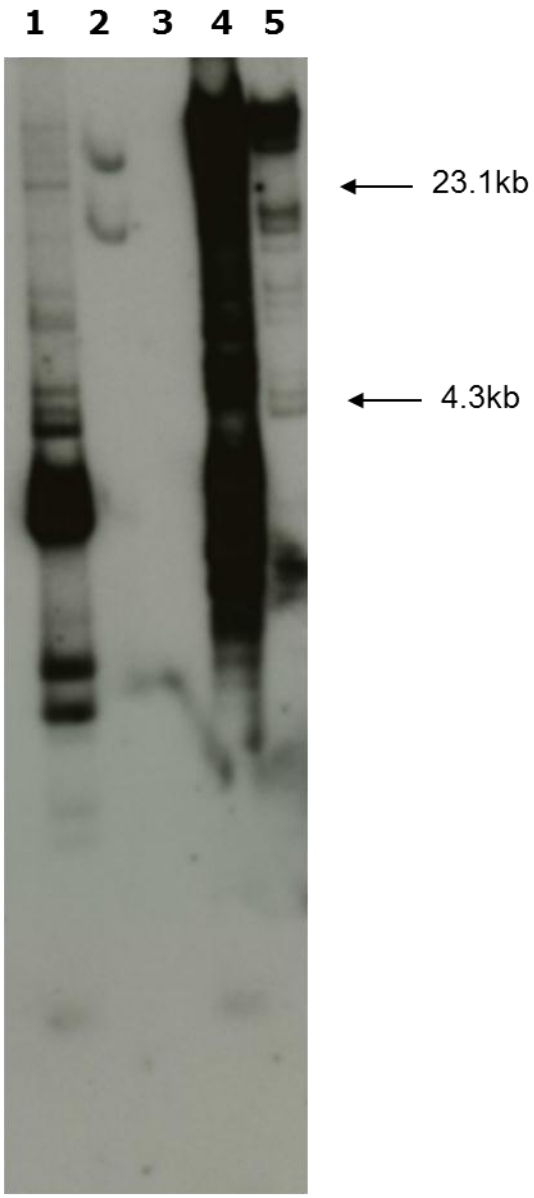


Figure 32: Autoradiogram of acrocentric blot with 380bpBstZ17I fragment. DNAs from various acrocentric chromosomes were digested with KpnI and run on a 1% agarose gel. The gel was then blotted and hybridized with the 380bpBstZ17I fragment at normal stringency (2XSSC, 60^{O}C). The lanes are 1. HC13, 2. HC14, 3. HC15, 4. HC21 and 5. HC22.

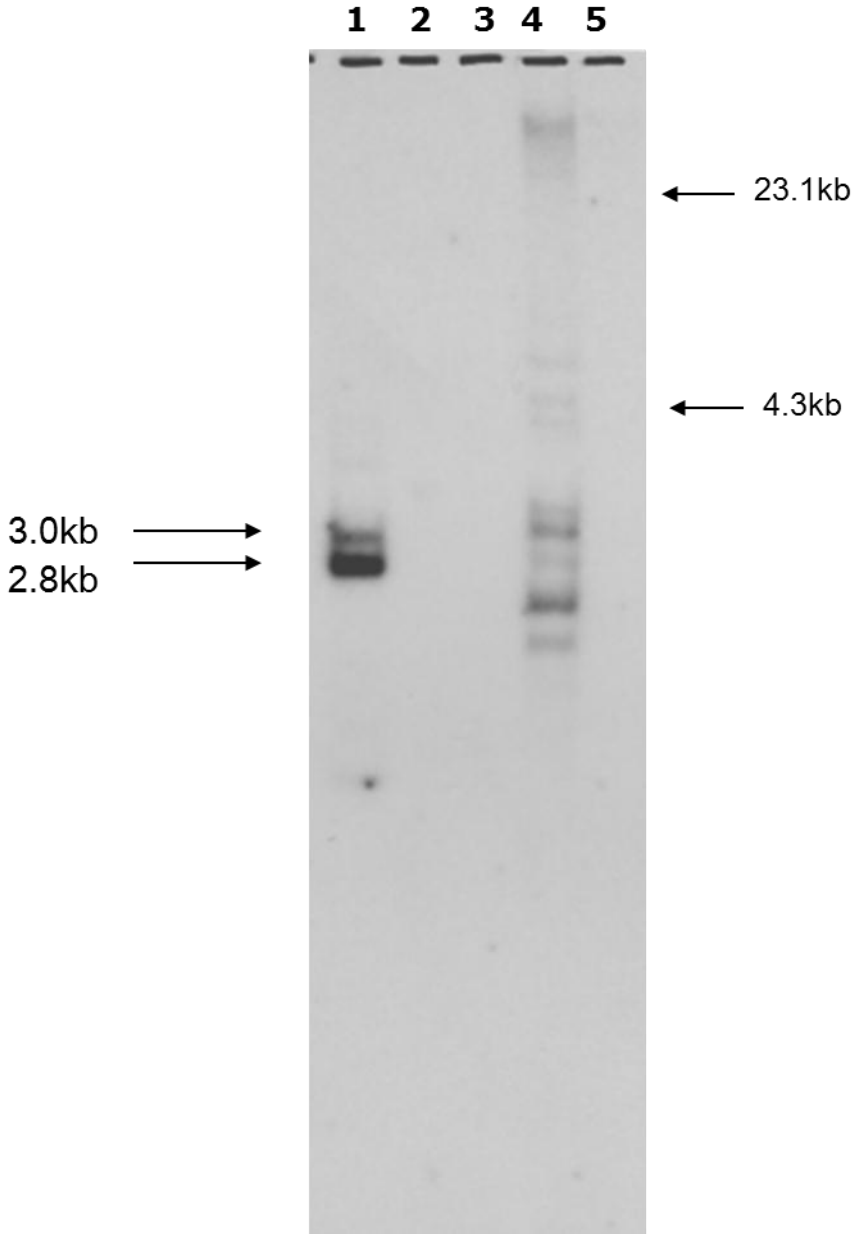


Figure 33: Autoradiogram of acrocentric blot with 380BstZ17I. DNAs from various acrocentric chromosomes were digested with KpnI and run on a 1% agarose gel. The gel was then blotted and hybridized with 380BstZ17I at high stringency (0.15XSSC at 60^{o}C). The lanes are 1. HC13, 2. HC14, 3. HC15, 4. HC21 and 5. HC22.

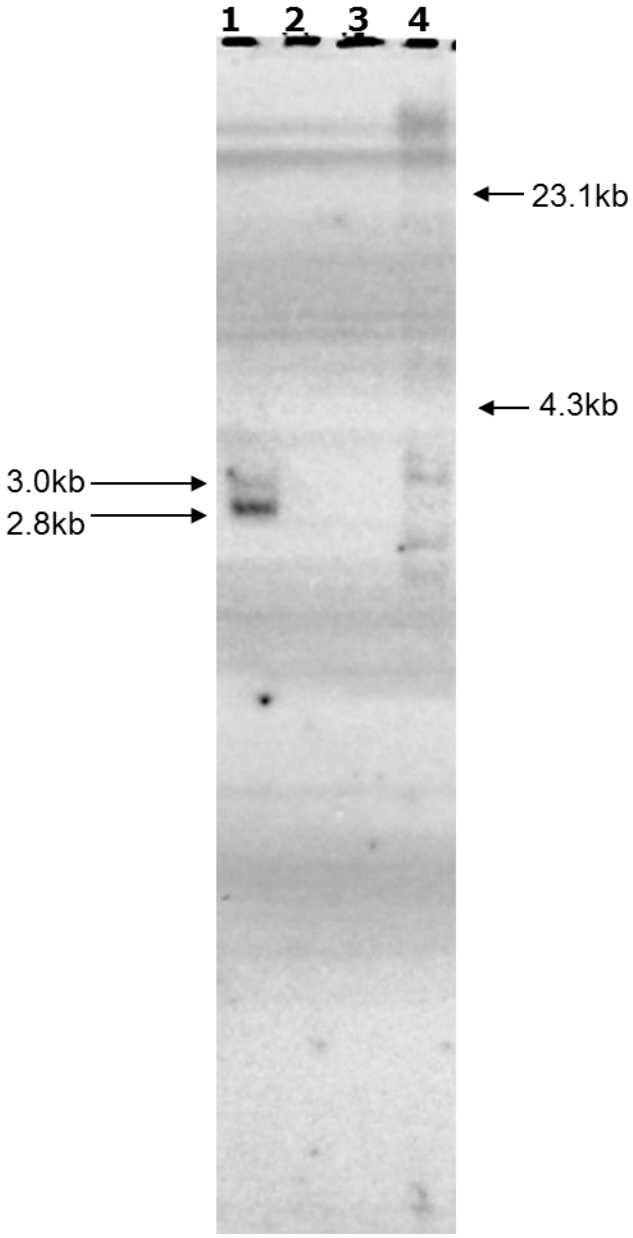


Figure 34: Autoradiogram of acrocentric blot with 380bpBstZ17I fragment. DNAs from various acrocentric chromosomes were digested with KpnI and run on a 1% agarose gel. The gel was then blotted and hybridized with 380BstZ17I at high stringency (0.15XSSC at 62.5^0C). The lanes are 1. HC13, 2. HC14, 3. HC15, 4. HC21.

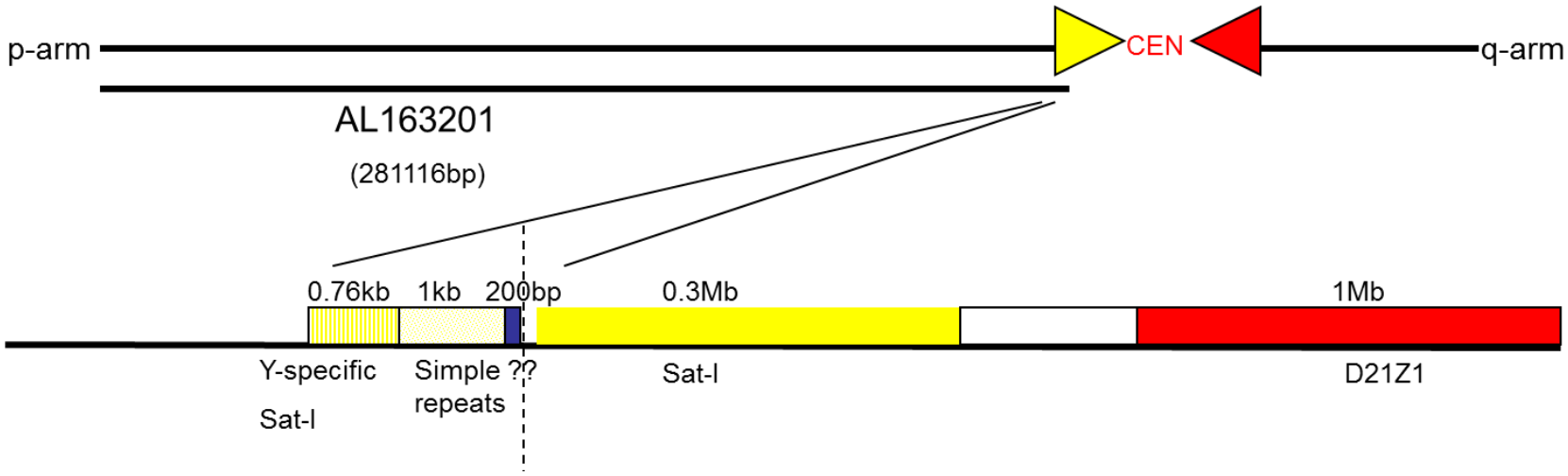


Figure 35: Initial analysis of Y-satI on HC21. AL162301 is shown. The q-arm end of AL163201 is expanded to show the approximate location and size of Y-specific sat I.

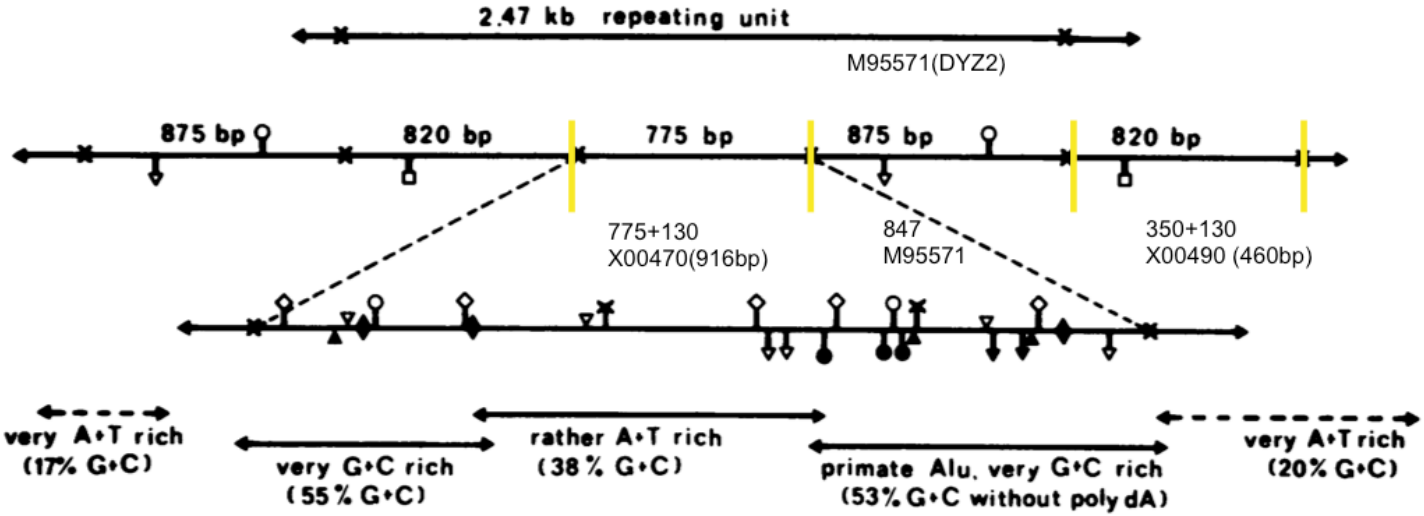


Figure 36: 2.47kb Y-satI repeating unit is shown. (FromVincent et al., 1984). It is composed of 3 *HinfI* fragments (shown as yellow lines) Accession numbers - X00470, M95571 and X00490. Prosser et al. sequenced 775bp of X00470 in addition to 130bp of M95571. They also sequenced 320bp X00490 and 130 of end part of M95571. The rest of the sequence of M95571 comes from Fenton et al. (1992).

bp # from 2.0Kb sequence on HC21	bp# from 2.4kb*SpeI* clone on HCY	BLAST match (%)
1-404	2012-2415	82
415-872	2-462	90
1731-2080	1731-2243	81
Alu match: 574-864	162-436	89
887-1759 (all ~100bp matches)	688-1920	76-88

Table 11: BLAST analysis results of 2.0kb sequence on HC21 against 2.4kbSpeI clone from HCY. The color coding matches Figure 37.

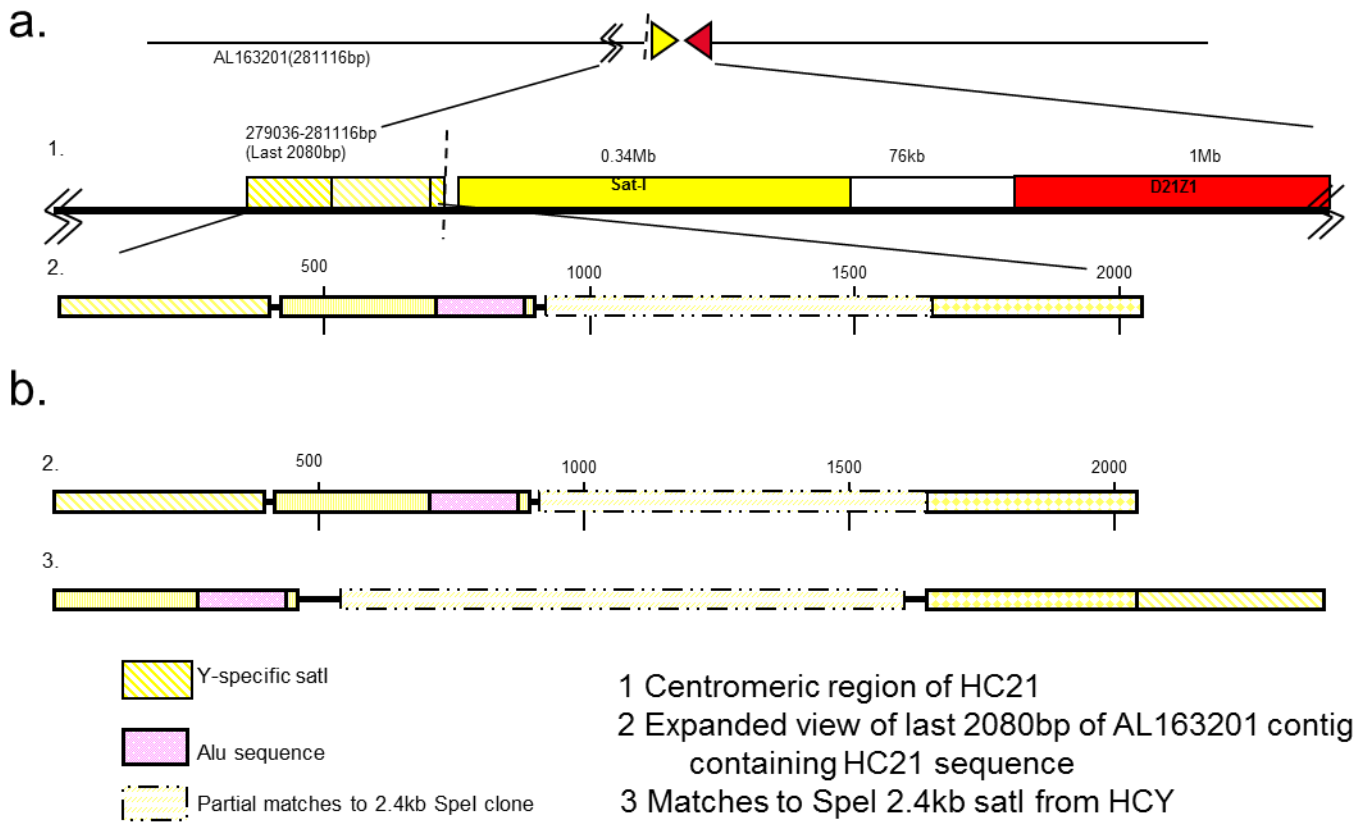


Figure 37 a. Organization of the p-arm end of HC21 centromere. 1. shows the centromeric map of HC21. 2. shows the expanded view of the Y-SatI region of HC21.

b. Comparison of Y-SatI region of HC21 and 2.4kb SpeI clone 2. shows the expanded view of the Y-SatI region of HC21. 3. Shows 2.4Kb SpeI match to HC21Y-satI. 2 and 3 show the organization and %match between the 2 sequences as indicated in Table 10.

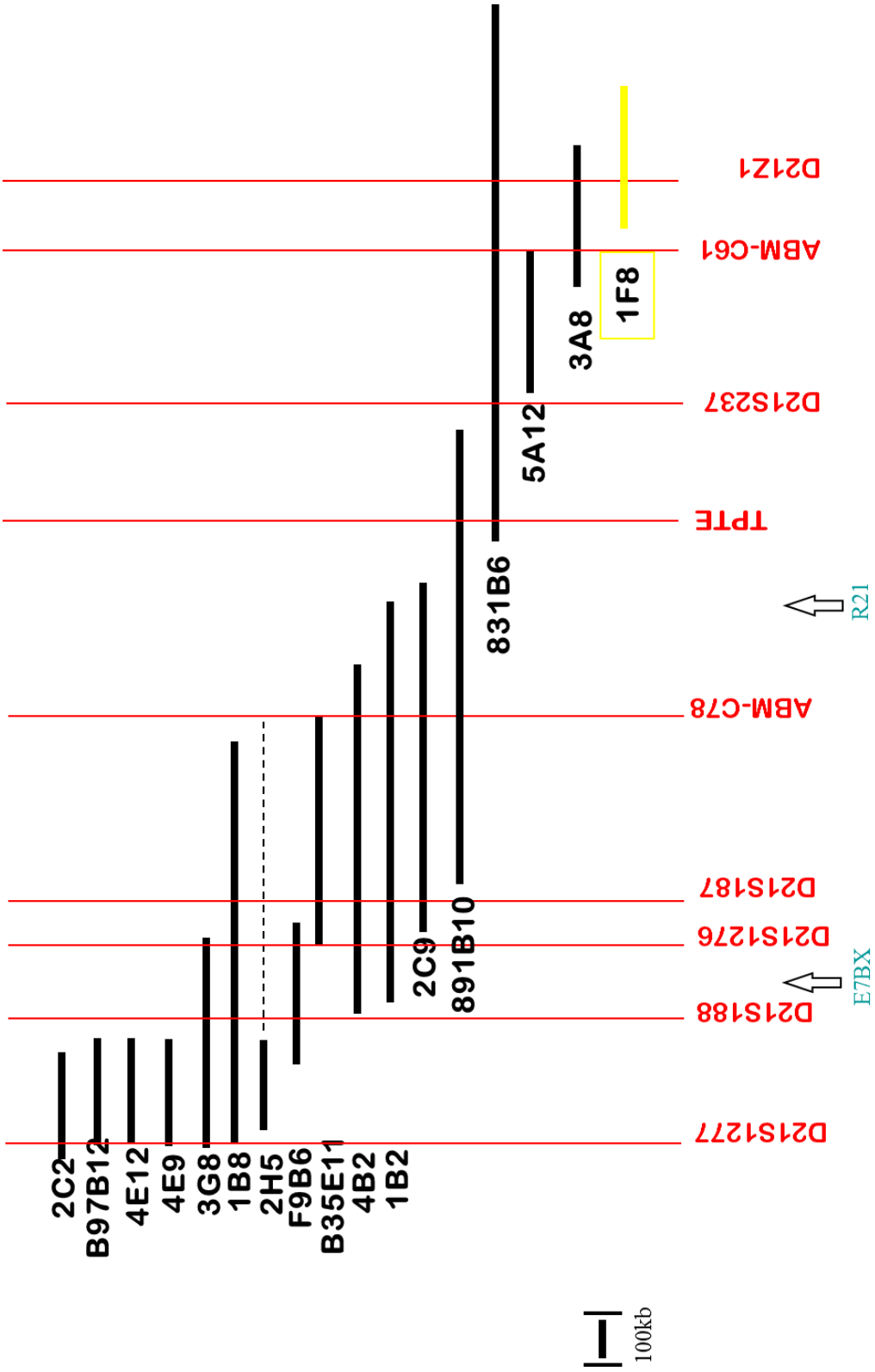


Figure 38: HC21 YAC map. The location of YAC1F8 is shown. This YAC clone spans the centromeric satellite I array and D21Z1 p-arm junction region and hence is used to determine the linkage between these centromeric clusters.

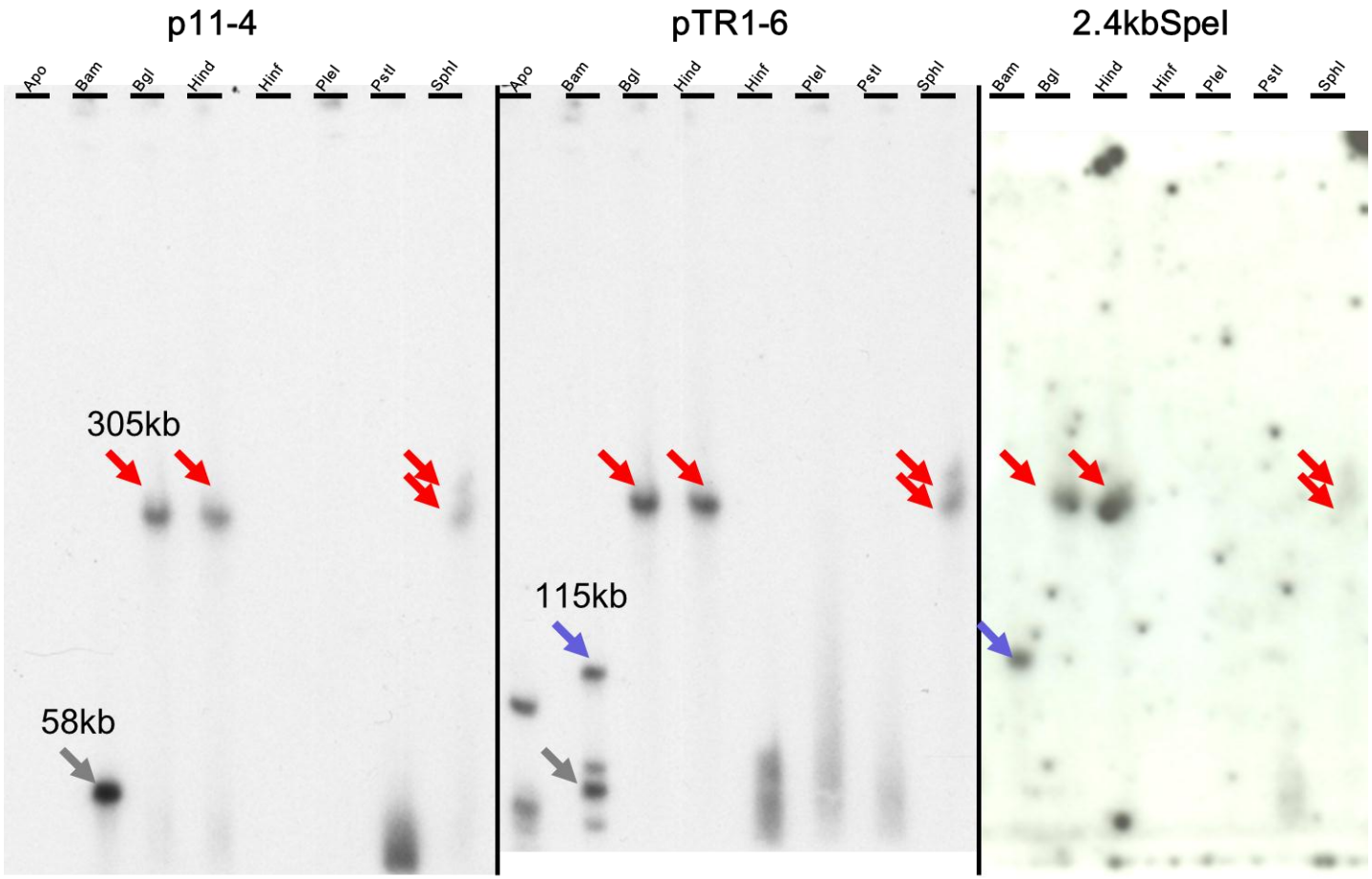


Figure 39: Mapping of YAC1F8. 1F8 was digested with various restriction enzymes as labeled on the autoradiograms and was then blotted onto a membrane. This membrane was then hybridized to p11-4(alphoid probe), pTR1-6(satI probe) and 2.4kbSpeI clone(Y-satI probe) respectively at normal stringency. The three bands that are present in on all three autoradiograms are marked with red arrows. The one band common to p11-4 and pTR1-6 is shown with green arrow. The one band common to pTR1-6 and 2.4kb *SpeI* is shown with blue arrow.

	p11-4	pTR1-6	2.4kb SpeI
ApoI	----	82.5, 27.5, 23.25kb	----
BamHI	58kb	115, 58,35,26.75kb	115kb
BglII	305kb	305kb	305kb
HindIII	305kb	305kb	305kb
SphI	343.5, 305kb	343.5, 305kb	343.5, 305kb

Table12: Size calibration of YAC1F8 fragments. Fragment sizes of YAC1F8 digested with different restriction enzymes are shown. The sizes are in Kb.

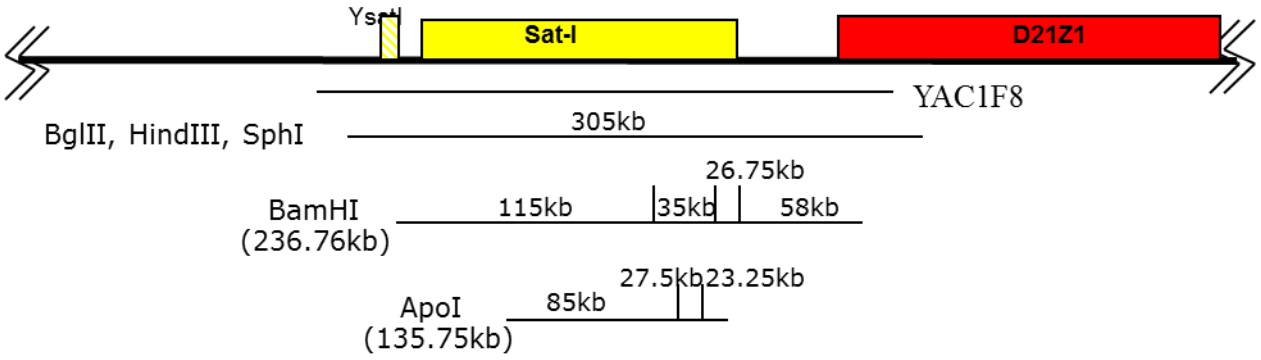


Figure 40: Map showing the organization of the centromeric clusters. The fragment sizes showing the organization of the centromeric cluster are shown below the map with the respective fragment sizes. The gap between D21Z1 and satI was previously determined to be 76kb which is now shown to be no more than 45kb.

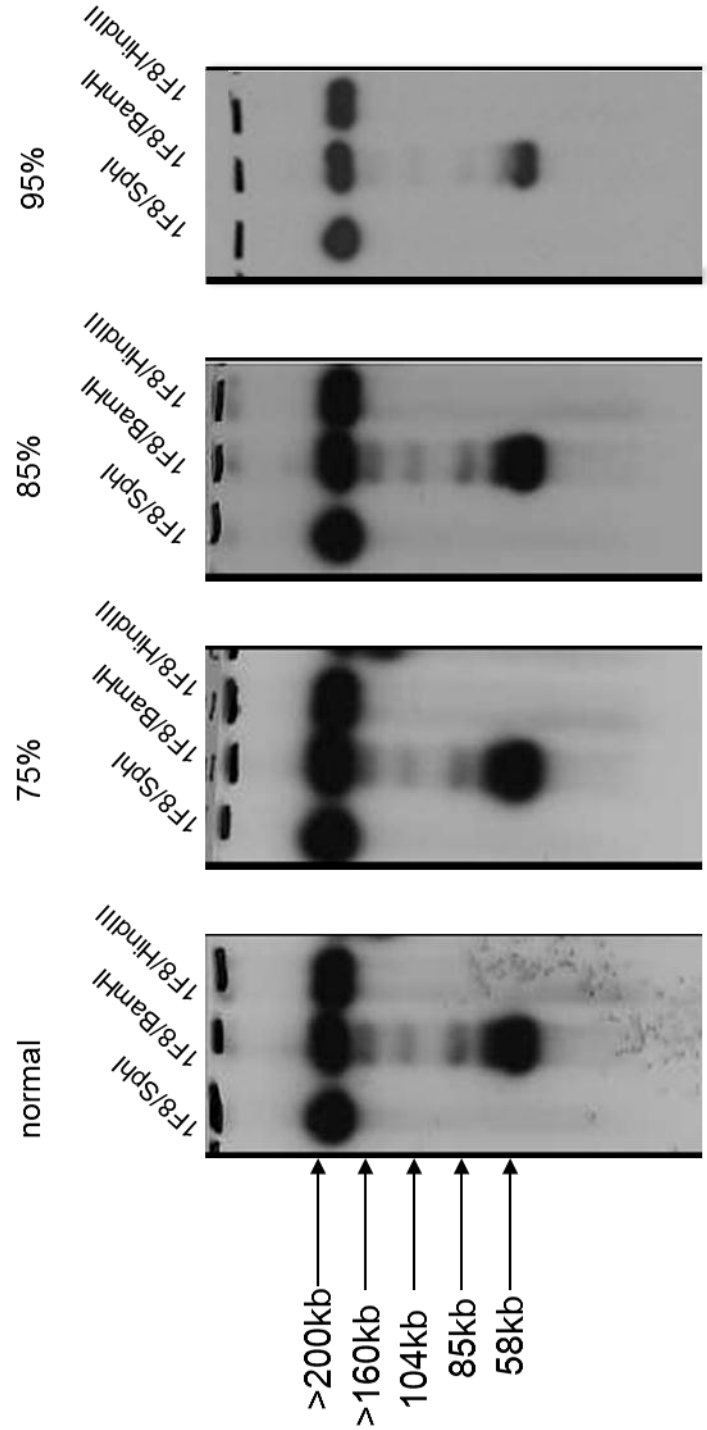


Figure 41: Autoradiogram of p11-4 hybridized at different stringencies. Hybridization of p11-4 at 70%, 75%, 85% and 95% are shown. The intensity of internal control band(>200kb) is comparable at all stringencies. There is no change in intensity in the 58kb junction region fragment.

CHAPTER FIVE

DISCUSSION

The centromeres have a vital role to play during cell division, functioning in sister chromatid adhesion, kinetochore formation, pairing of homologous chromosomes and the control of gene expression (reviewed in Larin and Mejia, 2002; Henikoff et al., 2001; Koch, 2000). Since centromeres are composed of heterochromatic tandem repeats, these regions were not included in the human genome project since it is hard to form a physical map and sequence large clusters of repeats. Some of the major clusters of repetitive DNA found in centromeres and other heterochromatic regions of the genome include alphoid DNA, beta satellite, satellites I, II and III, 48 bp repeats and interspersed repeats like LINES and SINES (Lee, et al., 1997).

The short (p) arms of the acrocentric chromosomes (13,14, 15, 21, 22) contain large amounts of heterochromatin and share substantial sequence similarities (Choo, 1990). Our lab is using the p-arm of chromosome 21 (HC21p) as a model for the structure of heterochromatin since it is the smallest chromosome with highest percent heterochromatin in the genome. We have shown that the structure of HC21p is a mosaic containing clusters of tandem repeats interspersed with low copy number repeats. The p-arm is ~20Mb in length, and many of the repetitive sequence families are at multiple locations both proximal and distal to the rDNA. There are several unexpected gene

sequences, like TPTE and BAGE, in this otherwise dense heterochromatic region. The map is still incomplete, containing some gaps (Figure 2).

Down syndrome arises due to an extra copy of chromosome 21 (trisomy 21). This occurs from improper segregation of chromosomes (non-disjunction) in meiosis resulting in gametes with an extra copy of HC 21. In order to study nondisjunction, it is necessary to be able to accurately determine the parent of origin and stage of meiosis in which the aberrant chromosomal segregation occurred. This is currently done using polymorphic markers on the long arm of human chromosome 21 (Hattori et al., 2000). Since these markers are located distal to the centromeric region, they can undergo recombination relative to the centromere. Thus, they can yield inaccurate data about the parent of origin and stage of meiosis. It would thus be advantageous to have an HC21 specific centromeric marker.

The centromeric satI cluster on HC21 was considered a good candidate for such a marker since such a cluster is not found in the centromere of any other acrocentric chromosome and it is polymorphic (Trowell et al., 1993; Roy et al., 2000). There is at least one other satI cluster on HC21p, but it is found at the distal end of the short arm and consists of a very different sequence subfamily than that in the centromeric satI cluster (Roy et al., 2000). Previous work showed that this proximal satI cluster contained substantial sequence heterogeneity (Patel, 2005). A region from within this satI cluster was previously subcloned from an HC21 YAC (Roy et al., 2000; Patel, 2005).

My first goal was to complete the sequence of this clone (pS21SatB) and test whether it could serve as an HC21-specific polymorphic centromeric marker. The

pS21SatB clone was completely sequenced and has a size of 4.8 kb. Various sub fragments of this clone were tested for specificity to HC21, but all of them showed the sequence to be present on HC13 as well as HC21. Although I showed that this sequence cannot serve as an HC21 specific centromeric marker, pS21SatB is the first satI clone obtained from chromosome 21 that is completely sequenced and whose organization is studied in detail. My work also characterized the heterogeneity of satI on HC21 and provided the first evidence of multiple satI families that are found on both HC13 and HC21.

The second goal was to study the organization of the p-arm junction region of the centromeric alphoid array and compare it to the previously determined q-arm junction region (Bozovsky et al., 2004). I found that the p-arm alphoid junction region is closely linked to the proximal satI cluster and does not have increased sequence heterogeneity as one gets closer to the end of the alphoid domain. This organization is unlike that of the q-arm junction region. I also mapped a Y-satI cluster on the HC21 p-arm and was able to complete a full physical map of the HC21 centromeric region.

Sequencing pS21SatB

Initially pS21SatB was sequenced using universal primers and was extended using primer walking. The total sequence generated by this approach was ~2.9 kb with single coverage (Figure 7). The sequence that was generated was mainly the ends of the insert (Patel et al., 2005). In order to complete the sequence and get double coverage, *EZ*::*TN*™ transposon reactions were carried out. Ideally transposons insert randomly within a clone, but in the case of pS21SatB they almost always inserted at the vector-

insert boundary. The reasons for such atypical transposon insertion behavior are unclear but suggest that the satI sequences are unfavorable insertion sites. Most of the sequences thus obtained from using the transposon reaction, gave double coverage for the ends of the clone (Figure 8). Primer walking was used to seal the gap and get double coverage for the remaining sequence. The problem with primer walking was the highly repetitive nature of this sequence, because the primer would bind to multiple sites within the sequence and hence create double peaks. Most of the primers gave double peaks at each nucleotide position, making it impossible to decipher the sequence. Part of the reason for getting double peaks was because the majority of the sequence was unknown initially and of low quality, making it difficult to do primer design. Thus the intensive internal sequence similarities lead to misleading and overlapping sequence runs when the primer-walking technique was used (Figures 11 and 12).

As a result of these problems, sub-cloning was used to finish sequencing the clone (Figure 15). Since the pS21SatB clone was quite big (4.8 kb satI + 4.1 kb vector) it posed a challenge to subclone smaller parts of the clone, because for subcloning there is a minimum DNA concentration requirement. This was difficult to obtain, as the subcloned fragments were no greater than 16% of the entire clone size. The problem was resolved once the entire satI clone was sub-cloned into pUC18, which is half the size of pSPORT (Figure 16). The chromatograms obtained from sequencing sub-cloned fragments were well resolved with no double peaks. This technique resulted in completing the sequence of pS21SatB unambiguously with double and sometimes quadruple coverage (Figure 15).

The size of the insert was determined to be 4.8 kb, and this clone was used for all further analyses.

My experience with this clone indicates that sequencing highly repeated tandem arrays should be done using a sub-cloning technique as it provides high quality sequence. Even though time consuming, it provides the best results for such difficult sequencing templates.

Sequence organization of pS21SatB

The pS21SatB sequence was compared to various known satI sequences from other chromosomes and also used to search the whole genome database. High %matches (>80%) were found to satI on HC13, HC4 and to a satI we previously identified at an unknown genomic position (Table 6). Three major matches to HC13 satI (pTR1-6) were found (Table 6). With these exceptions, there are no sequences in the database that are highly similar to pS21SatB (>85% and covering the entire sequence without any gaps). Over 400 matches were obtained with the whole genome BLAST. The highest percent similarity over the longest distance was found to satI sequences from HC13, HC16, HC8 and HC4 (Table 7). Most of the matches are small and have low % similarity to pS21SatB, indicating that satI is a highly diverged genomic sequence family. Thus pS21SatB is apparently a unique satI displaying more similarity to acrocentric satI from HC13 then to satI from other chromosomes. It might thus contain a putative HC21 centromere specific probe.

Except for HC16 and HC8, the satI sequences were on individual clones and were not found in the mapped regions of the genome, suggesting that there is still a lot of satI

sequence organization in the genome that is uncharacterized. Preliminary analysis of the HC8 BAC clone (Acc# AC022616) and NCBI Map Viewer analysis showed that the satI match found on this chromosome is ~0.5Mb from the centromeric alphoid cluster. This is a much greater distance apart than analogous sequences on HC21p (see below). The satI cluster on HC8 is monomeric, like that on HC21p, and so suggests that the HOR satI clusters are not features of centromeric regions.

Monomeric and Higher Order Satellite I Structures

The degree of conservation among the monomers is around 88% in pS21SatB, which explains why it was difficult to find unique primers for sequencing it initially. In contrast to the 42 bp monomers of HC13satI, the HC21satI studied here has a 41 bp monomer (Figure 20). This suggests that even though HC13 and HC21 satI might interact through non-homologous p-arm exchanges, there is a basic difference between the two, indicating derivation from different ancestral monomers.

pS21SatB does not contain a clear HOR, as evidenced by the dot plots (Figure 22) and internal sequence comparisons (Table 9). Southern blot results also show that the 3 kb HOR prominent on HC13 is not seen on HC21 (Figure 24). The sequence heterogeneity within the monomers indicates that pS21SatB is a monomeric satI sequence, and that it is a different satI family from pTR1-6 (Table 9). The highly heterogeneous sequence in pS21SatB could explain why pTR1-6 hybridizes to some but not all regions within the HC21 satI centromeric cluster (Trowell et al., 1993; Roy et al., 2000). The monomeric nature of the cluster would also account for its heterogeneous

sequence composition, since no correcting mechanisms would be in place, as can exist for HOR clusters.

The present work shows that there are at least two satI clusters on HC21, both of which are monomeric. In addition to the centromeric cluster represented by pS21SatB, there is another monomeric cluster on distal 21p, characterized by N6.4 (Roy et al., 2000), which is distinctly different from pS21SatB (Figure 20). HC13 has only one characterized satI cluster (Kalitsis et al., 1992) that is composed of an HOR (pTR1-6) clearly related to pS21SatB (Figure 20). Our lab's recent work has shown that HOR clusters can arise within monomeric sequences during primate evolution (Ennesser and Doering, unpublished). Thus, the pTR1-6 HOR could have arisen from monomeric regions related to pS21SatB. A future phylogenetic study of the satellite I sequence family would thus be of interest.

Identification of regions within pS21SatB putatively specific for HC21

Our preliminary analysis suggested that pS21SatB has regions specific for the HC21 centromere. The 1176 bpBstZ17I fragment was initially chosen because it matched the criteria used to pick a HC21specific marker sequence. However, blot results show that even at 95% stringency this sequence is present on HC13 more prominently than on HC21 (Figure 26). This result was reproducible with other copies of HC21 and 13, indicating this sat I organization is a conserved feature of these acrocentric p-arms (Figure 27). Because the preliminary results suggested that there were regions in the1176 bp BstZ17I fragment, putatively specific, with very low 83-85% matches to pTR1-6, the fragment was re-examined. Upon BLASTing pS21SatB vs. pTR1-6 it was found that

there were regions within the 1176 bp sequence that matched about 95-100% with pTR1-6 over a 100 bp stretch (Figure 28). This was identified as the problem and thus the criteria for picking a probe sequence from pS21SatB were modified. The length was significantly reduced to 300-700 bp and the % similarity to pTR1-6 was reduced to 95%. Shorter probe sequence-380BstZ17I was obtained that best matched the new criteria. Blots showed that this probe did not hybridize to the pTR1-6 sequence suggesting it might be HC21 specific (Figure 31). When this was used as a probe on the acrocentric cell panel at a low stringency, the intensity of hybridization was very high on HC21compared to HC13, as expected. However, as the stringency was raised, the bands on HC13 were still present, showing that even this shorter sequence was not specific to HC21 (Figures 32 and 33).

The results indicate that even though the pS21SatB-like sequence is not present in the centromeric region of any other acrocentric chromosome, including HC13 (Trowell et al., 1993), similar sequences must be present elsewhere on HC13. Thus, there may be monomeric satI clusters on HC13 in addition to the well-characterized HOR cluster of pTR1-6. My results (Figure 32) also show that there are satI clusters related to pS21SatB on HC14 and 22, but not on HC15, suggesting that there may be a variety of pS21SatB-like clusters throughout the genome. It would be interesting to check for the presence of these clusters on other chromosomes as well.

An HC21 specific centromeric marker sequence has not yet been obtained, though these experiments helped learn more about satI sequences in general, and I also showed the first evidence for the presence of pS21SatB-like sequences on HC13. This is the first

satI clone obtained from chromosome 21 that is completely sequenced and studied in detail.

Presence of Y-satI on HC21 and its linkage to centromeric D21Z1 and SatI cluster

De Sario et al. (2003) analyzed the genomic structure in the juxtacentromeric region of human chromosome 21. Based on their analysis, they suggested that the BAC contig containing the centromeric end of HC21p (AL163201) had been completely annotated and contained a 0.6 kb satI cluster with 80% nucleotide similarity to GenBank Accession Number X00470. My closer look at this region showed that the size of this satI cluster is actually 0.76 kb and the identity to X00470 is 82%. Analyzing the sequence of this region further showed there was actually a 1.2 kb domain following the 0.76 kb region that was not annotated or characterized. This region has one AluSc element followed by 1 kb of simple TA repeats and 200 bp that did not match satI or the X00470 sequence (Figure 35). Moreover BLAST analysis showed that this ~2.0 kb region at the end of the contig is in fact the complex Y-specific satI sequence, which is completely different from classic satellite I (Prosser et al., 1986; Figures 35 and 36). A comparison in the organization of YsatI between HC21 and HCY can be seen in Figure 37. Based on these results it is evident that they are similar in organization, containing the AluSc element.

Babcock et al. (2007) characterized Y-specific satI as a 2.4 kb repeat that is ubiquitous on Yq and found in multiple copies in the pericentromeric region of 22q. They suggested the sequence was also in the pericentromeric p-arm of all human acrocentric chromosomes. My work shows that there is only one copy of the sequence in

pericentromeric 21p (Figure 35), and that it is poorly conserved relative to copies on Yq (Figure 37 and Table 11). YsatI is interspersed with satII clusters on Yq (Babcock et al., 2007) while that organization is not seen on 21p. Thus YsatI has evolved differently on the acrocentric chromosomes and does not even have the same organization on all acrocentrics.

The YsatI clusters on 22q11.2 are associated with various genomic translocation disorders (Babcock et al., 2007). Further work should determine whether copies of the sequence on other acrocentric chromosomes are also involved in translocations.

The current work not only helped identify a novel Y-satI cluster in the centromeric region of HC21 but also updated the centromeric map of HC21 by establishing clear linkage relations between D21Z1, satI and Y-SatI (Figure 40). The size of the centromeric satellite I cluster was previously determined to be 0.20Mb (Trowell et al., 1993; So et al., 1997), and my work shows the cluster to be about the same size (Figure 40). The YACs I studied are derived from a different population than that used in the earlier studies, indicating a significant cluster size polymorphism in the population, not unlike that seen for the D21Z1alphoid cluster (Maratou et al., 2000). While the distance between D21Z1 and satI was previously estimated at 76 kb (Shukair et al., 2005), my map places this distance at no greater than ~50 kb (Figure 40). Trowell et al. (1993) and So et al (1997), placed the satI cluster on the q-arm side of the alphoid cluster, but my map clearly shows satI to be on the p-arm side and proximal to the Y-satI sequence (Figure 40). This location of satI on 21p confirms more recent mapping (Guipponi et al., 2000).

Maratou et al. (2000) indicated that small D21Z1 alphoid cluster size is a risk factor for meiosis I non-disjunction in DS. The study used a *BamHI* fragment that my map indicates includes satellite I as well as centromeric D21Z1 alphoid DNA (Figure 5). Thus when the size of the alphoid cluster was estimated in previous work, it would include the size of the satellite I present in the *BamHI* fragment. Thus the basis for the size variation could be the alphoid cluster, the satI cluster or a combination of the two. If further mapping allows D21Z1 and satI to be separated in distinct fragments, it would be interesting to use pS21SatB as a probe to determine the size variation of centromeric satI clusters in DS patients. Since HC21 is the only acrocentric centromere where alphoid and satI are linked (Trowell et al., 1993), fragments containing both D21Z1 and the satI cluster could then be analyzed for variation in the alphoid cluster size independent of the highly similar alphoid cluster on HC13 (Wang et al., 1999).

Determining the short arm junction region of D21Z1

Determining the organization of the short arm junction region of D21Z1 allows us to test the out-of-register recombination model that predicts that there is gradual loss of homogeneity at both the ends of a repetitive sequence cluster (Smith, 1976; Willard et al., 2006). Previous study shows that the q-arm junction region of D21Z1 does not have such a gradual loss (Bozovsky et al., 2004), and the present study shows that the p-arm junction region of D21Z1 also does not show any gradual loss of homogeneity (Figure 41). This is the first study where both end junctions of the same centromeric alphoid cluster have been studied in detail to determine the validity of this model. The combined results of the current study and the work done by Bosovsky et al., (2004) suggest that

HC21 junction regions contradict the prediction of the unequal crossover model. A different model is thus needed to explain the evolution of this region of HC21. Bosovsky et al., (2004) suggested that the higher order D21Z1 array might have been inserted in an evolutionarily older monomeric array. This study provides evidence in support of that suggestion as both –p-arm and q-arm junction regions show no gradual loss of homogeneity.

Characterization of the D21Z1 junction regions is now complete. My work has linked the HC21p-arm map with the full HC21q-arm map. This map of HC21 can be used as a model for the structure of all acrocentric chromosomes and particularly in understanding the function and evolution of the centromere in more detail.

Further work

pTR1-6 is an HOR satI family that has been commonly used for studying the presence of all satI on human chromosomes (Kalitsis et al., 1992). pS21SatB is now the first well-characterized monomeric satI sequence. It would be interesting to see the results when it is used as a probe to identify similar sequences elsewhere in the genome. Also a phylogenetic study of this sequence would help understand the evolution of satI sequences in general.

pS21SatB can also be used as a probe to find other variants on the HC21 centromeric satI cluster that might be chromosome specific. Weak matches to pTR1-6 and p21SatB might be good candidates for HC21 centromeric markers. YsatI cluster on HC21 might also be a good candidate for being an HC21-specific centromeric marker,

since preliminary *in silico* analysis shows poor % matches to YsatI from HCY. The current work has clarified the parameters required for future attempts at finding an HC21-centromere specific marker *in silico* and experimentally. The results of this study and the satI clone can also be used to study the role of satI in centromere function.

WORKS CITED

Ando, S., Yang H., Nozaki N., Okazaki T., and Yoda K. 2002. CENP-A, -B, and -C chromatin complex that contains the I-type -satellite array constitutes the prekinetochore in HeLa cells. Mol. Cell. Biol. 22: 2229–2241.

Antonarakis SE, Lyle R, Dermitzakis ET, Reymond A, Deutsch S. 2004 Chromosome 21 and Down syndrome: From genomics to pathophysiology. Nat Rev Genet. 5: 725-38. Review.

Babcock M., Yatsenko S., Stankiewicz P., Lupski J.R. and Morrow B.E. 2007. AT-rich repeats associated with chromosome 22q11.2 rearrangement disorders shape human genome architecture on Yq12. Genome Res. 17: 451-460.

Bjerling P., Ekwall K. Centromere domain organization and histone modifications 2002. Brazilian Journal of Medical and Biological Research 35: 499-507

Bozovsky, M.R., S.A. Shukair, M.R. Cummings and J.L. Doering. 2004. Organization of the regions flanking the centromere of human chromosome 21. [abstract 1567]. Available from http://www.ashg.org/genetics/ashg04s/

Chakravarti, A. and Lynn, A. 1999. Meiotic mapping in humans. In genome analysis: A laboratory manual, Vol. 4. Mapping genomes (ed. B. Birren, et al.,), pp. 1–69. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Choo, K.H. 1990. Role of acrocentric cen-pter satellite DNA in Robertsonian translocations and chromosomal nondisjunction. Mol. Biol. Med. 7: 437-449

Choo, K.H.A., 1997. The centromere, Oxford University Press, 1-304

Cooper, K.F., Fisher, R.B., Tyler-Smith, C. 1993. Structure of the sequences adjacent to the centromeric alphoid satellite DNA array on the human Y chromosome. J. Mol. Biol. 230: 787-799.

Edelmann, L., Spiteri, E., Koren, K., Pulijaal, V., Bialer, M.G., Shanske, A., Goldberg, R., and Morrow, B.E. 2001. AT-rich palindromes mediate the constitutional t(11;22) translocation. Am. J. Hum. Genet. 68: 1–13.

Eichler, E.E., Clark, R.A., and She, X. 2004. An assessment of the sequence gaps: unfinished business in a finished human genome. Nature Reviews: 5: 348-355

Gosden, J.R., Lawrie, S.S., Gosden, C.M. 1981. Satellite DNA sequences in the human acrocentric chromosomes: information from translocations and heteromorphisms. Am J Hum Genet 33: 243-251

Grimes, B. and Cooke, H. 1998. Engineering mammalian chromosomes. Hum. Molec. Genet. 7: 1635-1640

Hassold T. and Chiu D.1985. Maternal age-specific rates of numerical chromosome abnormalities with special reference to trisomy. Hum. Genet. Vol. 70: 11-17

Hassold, T. and Sherman, S. 2000. Down syndrome: genetic recombination and the origin of the extra chromosome 21. Clin Genet. 57: 95-100.

Hattori, M., A. Fujiyama, T. D. Taylor, H. Watanabe, T. Yada, H. S. Park, A. Toyoda, K. Ishii, Y. Totoki, D. K. Choi, Y. Groner, E. Soeda, M. Ohki, T. Takagi, Y. Sakaki, S. Taudien, K. Blechschmidt, A. Polley, U. Menzel, J. Delabar, K. Kumpf, R. Lehmann, D. Patterson, K. Reichwald, A. Rump, M. Schillhabel, A. Schudy, W. Zimmermann, A. Rosenthal, J. Kudoh, K. Schibuya, K. Kawasaki, S. Asakawa, A. Shintani, T. Sasaki, K. Nagamine, S. Mitsuyama, S. E. Antonarakis, S. Minoshima, N. Shimizu, G. Nordsiek, K. Hornischer, P. Brant, M. Scharfe, O. Schon, A. Desario, J. Reichelt, G. Kauer, H. Blocker, J. Ramser, A. Beck, S. Klages, S. Hennig, L. Riesselmann, E. Dagand, T. Haaf, S. Wehrmeyer, K. Borzym, K. Gardiner, D. Nizetic, F. Francis, H. Lehrach, R. Reinhardt and M. L. Yaspo (2000). The DNA sequence of human chromosome 21. Nature 405: 311-9.

Henikoff, S., K. Ahmad, and H.S. Malik. 2001. The centromere paradox: stable inheritance with rapidly evolving DNA. Science. 293: 1098–1102

Hettinger, R., J. Bavarian, P. Patel, T. George, M. Puckelwartz, S. McCutcheon, M. Cummings and J. Doering. 2005. Organization of satellite I and satellite III sequences on human chromosome 21. [abstract 1348]. Available from http://www.ashg.org/genetics/ashg05s/

Hernandez, D. and Fisher, E.M.C. 1996. Down syndrome genetics: unraveling a multifactorial disorder. Hum. Molec. Genet. 3: 398-403

Horvath, J.E, Viggiano L., Loftus, B.J., Adams, M.D., Archidiacono, N., Rocchi, M. and Eichler, E.E. 2000. Molecular structure and evolution of an alpha satellite/non-alpha satellite junction at 16p11. Hum Mol Genet. 9: 113:123

Horvath J.E., Bailey J.A., Locke D.P. and Eichler E.E. 2001. Lessons from the human genome: transitions between euchromatin and heterochromatin. Hum Mol Genet., 10: 2215-2223

Ikeno, M., Masumoto, H., Okazaki, T. 1994. Distribution of CENP-B boxes reflected in CREST centromeric antigenic sites on long range -satellite DNA arrays of human chromosome 21. Hum. Mol. Genet. 3: 1245-1257.

Kalitsis, P. Earle, E., Vissle B., Shaffer L. G., Choo K. H. 1993. A chromosome 13-specific human satellite I DNA subfamily with minor presence on chromosome 21: Further studies on Robertsonian translocations. Genomics 16: 104-122

Kehrer-Sawatzki, H., Haussler, J., Krone, W., Bode, H., Jenne, D.E., Mehnert, K.U., Tummers, U and Assum, G., 1997. The second case of a t(17;22) in a family with neurofibromatosis type 1: Sequence analysis of the breakpoint regions. Hum. Genet. 99: 237–247.

Kurahashi, H., Shaikh, T.H., Hu, P., Roe, B.A., Emanuel, B.S., and Budarf, M.L. 2000. Regions of genomic instability on 22q11 and 11q23 as the etiology for the recurrent constitutional t(11;22). Hum. Mol. Genet. 9: 1665–1670.

Koch, J. 2000. Neocentromeres and alpha satellite: A proposed structural code for functional human centromere DNA. Hum. Mol. Genet. 9: 149–154.

Kurnit D.M., Stewart G.D., Hassold T.J., Berg A., Watkins P., Tanzi R. 1998. Trisomy 21 (Down syndrome): studying nondisjunction and meiotic recombination by using cytogenetic and molecular polymorphisms that span chromosome 21. Am J Hum Genet. 42: 227–236.

Lamb N.E., Freeman S.B., Savage-Austin A., Pettay D., Taft L., Hersey J., Gu Y., Shen J., Saker D., May K.M., Avramopoulos D., Petersen M.B., Hallberg A., Mikkelsen M., Hassold T.J., Sherman S.L.1996. Susceptible chiasmate configurations of chromosome 21 predispose to non-disjunction in both maternal meiosis I and meiosis II. Nature Genet. 14: 400-405.

Lamb N.E, Yu K, Shaffer J, Feingold E, and Sherman S. 2005. Association between Maternal Age and Meiotic Recombination for Trisomy 21. Am. J. Hum. Genet. 76: 91–99

Larin, Z. and Mejia, J. 2002. Advances in human artificial chromosome technology. Trends Genet. 18: 313-319.

Laurent A.M, Li M, Sherman S, Roizès G and Buard J. 2003. Recombination across the centromere of disjoined and non-disjoined chromosome 21. Human Molecular Genetics. 12 2229-2239

Lee, C., Wevrick, R., Fisher, R.B., Ferguson, M.A., Lin, C.C. 1997. Human centromeric DNAs. Hum. Genet. 100: 291-304.

Lo, A., Liao, G., Rocchi, M., and Choo K.H. 1999. Extreme reduction of chromosome-specific alpha-satellite array is unusually common in human chromosome 21. Genome Research. 9: 895-908

Manuelidis, L. 1978. Chromosomal localization of complex and simple repeated human DNAs. Chromosoma. 66: 23-32.

Mantovani B., Plohl M, Luchetti A, Mestrović N, 2007. Satellite DNAs between selfishness and functionality: Structure, genomics and evolution of tandem repeats in centromeric (hetero)chromatin Gene. 409: 72-82.

Mashkova, T.D., Tyumeneva, I.G., Zinovileva, O.L., Romanova, L.Y., Jabs, E., and Aleksandrov, I.A.1996. Centromeric alpha satellite DNA at euchromatin/ heterochromatin boundary of human chromosome 21. Mol Biol. 30: 617-625

Masumoto H., Ohzeki JI., Nakano M., Okada T. 2002. CENP-B box is required for de novo centromere chromatin assembly on human alphoid DNA. The Journal of Cell Bio. 159: 765-775

Maratou, K., Siddique, Y., Kessling, A.M. and Davis, G.E. 2000. Variation in alphoid DNA size and trisomy 21: a possible cuase of nondisjunction. Hum Genet. 106: 525-530

Meyne, J., Goodwin, E.H., and Moyzis R.K. 1993. Chromosome localization and orientation of the simple sequence repeat of human satellite I DNA. Chromosoma 103: 99-103

Musich P.R., Dykes R.J. 1986. A long interspersed (LINE) DNA exhibiting polymorphic patterns in human genomes. Proc. Natl. Acad. Sci. Vol. 83: 4854-4858

Politi, V., G. Perini, S. Trazzi, A. Pliss, I. Raska, W.C. Earnshaw, and G.D. Valle. 2002. CENP-C binds the -satellite DNA in vivo at specific centromere domains. J. Cell Sci. 115: 2317–2327

Prosser J., Frommer M., Paul C., Vincent P.C. 1986. Sequence relationships of three human satellite DNAs. J. Molecular Biology. 187: 145 – 155.

Reed K.C., Mann D.A. Rapid transfer of DNA from agarose gels to nylon membranes. Nucleic Acids Research, 1985, Vol. 13: 7207-7221

Roy, M., J. So, M. Cummings and J. Doering. 2000. Organization of satellite I on human chromosome 21. Am. J. Hum. Genet. 67 (Suppl.): 1473.

Schueler, M. G., A. W. Higgins, M. K. Rudd, K. Gustashaw and H. F. Willard. 2001. Genomic and genetic definition of a functional human centromere. Science 294: 109-15.

Schueler, M.G., Higgins A. W., Rudd, M.K., Gustashaw, K. and Willard, H.F. 2001. Genomic and genetic definition of a functional human centromere. Science 294: 109-115

Schueler M.G., and Sullivan B.A. 2006. Structural and functional dynamics of human centromeric chromatin. Annu. Rev. Genom. Human Genet. 7: 301-313.

Schmid M., Guttenback M, Nanda I., Studer R, Epplen JT., 1990. Organizatin of DYZ2 repetitive DNA on the human Y chromosome. Genomics 6: 212-218

Sherman, S.L., Peterson, M.B., Freeman, S.B., Hersey, J., Pettay, D., Taft, L., Frantzen, M., Mikkelsen, M., Hassold, T.J. 1994. Non-disjunction of chromosome 21 in maternal meiosis I: evidence for a maternal age-dependent mechanism involving reduced recombination. Hum. Mol. Genet. 3: 1529-1535.

Smith, G. 1976. Evolution of repeated DNA sequences by unequal crossover. Science. 191: 528-535.

So, J, Stevens, C, Weise, I, Cummings, M, Doering, JL. 1997. "Detailed physical mapping of the centromere region of human chromosome 21." Am J Hum Genet. 61 (suppl.):2298.

Stewart G.D., Hassold T.J., Berg A, Watkins P, Tanzi P, and Kurnit D.M. 1988. Trisomy 21 (Down syndrome): studying nondisjunction and meiotic recombination by using cytogenetic and molecular polymorphisms that span chromosome 21. Am. J. Hum. Genet. 42: 227-236

Tagarro, I., Wiegant, J., Raap, A.K., Gonzalez-Aguilera, J.J. and Fernandez-Peralta, A.M. 1993. Assignment of human satellite 1 DNA as revealed by fluorescent in situ hybridization with oligonucloetides. Human Genetics 93: 125-128

Trowell, H.E., Nagy, A., Vissel, B. & Choo, K.H.A. 1993. Long-range analyses of the centromeric regions of human chromosomes 13, 14 and 21: identification of a

narrow domain containing two key centromeric DNA elements. Hum. Mol. Genet.2: 1639−1649

Tyler-Smith, C., Oakley, R.J., Larin, Z., Fisher, R.B., Crocker, M., Affara, N.A., Fergusen-Smith, M.A., Zuffardi, O., Jobling, M.A. 1993. Localization of DNA sequences required for human centromere function through an analysis of rearranged Y chromosomes. Nature Genet. 5: 368-375.

Tyler-Smith, C., Willard, H.F. 1993. Mammalian chromosome structure. Curr. Opinion in Genetics and Development 3: 390-397.

Ventura, M., Archidiacono, N. and Rocchi, M., 2001. Centromere emergence in evolution. Genome Res. 11: 595–599

Vincent P.C., Frommer M., Prosser J., 1984. Human satellite I sequences include a male specific 2.47 kb tandemly repeated unit containing one Alu family member per repeat. Nuc. Acids Res. 12: 2887-2900

Vissel, B. and Choo K.H. 1991. Four distinct alpha satellite subfamilies shared by human chromosomes 13, 14 and 21. Nuc. Acids Res. 19: 271-277

Willard H.F., Waye J.S. 1987. Hierarchical order in chromosome-specific human alpha satellite DNA. Trends Genet 3: 192-198

Willard H.F., Warburton P.E, Waye J.S. 1993. Nonrandom localization of recombination events in human alpha satellite repeat unit variants: implications for higher-order structural characteristics within centromeric heterochromatin. Mol Cell Biol. 13: 6520-6529

Willard H.F. 2001. Neocentromeres and human artificial chromosomes: An unnatural act. PNAS. 98(10): 5374-5376

Willard H.F. Rudd M.K., Wray G.A. 2006. The evolutionary dynamics of α-satellite. Genome Res. 16: 88–96

Wevrick, R., Wilaard, V.P. and Willard, H.F. 1992. Structure of DNA near long tandem arrays of alpha satellite DNA at the centromere of human chromosome 7. Genomics 14: 912-923

Wang S.Y.,Cruts M., Del-Favero J.,Yi Zhang, Tissir F., Potier M.C., Patterson D., Nizetic D., Bosch A., Chen H., Bennett L., Estivill X., Kessling A., Antonarakis S.E., and Broeckhoven C.V., 1999. A high-resolution physical map of human chromosome 21p using yeast artificial chromosomes. Genome Research 9: 1059–1073

Wong, L.H. and Choo, K.H.A., 2001. Centromere on the move. Genome Res. 11: 51

VITA

Riddhi Patel was born in 1981 in Ahmedabad, India. She graduated from the Gujarat Arts and Science College with a Bachelor's degree in Microbiology in 2002. She then did her post-graduation from the Gujarat University School of Sciences and obtained a Master's degree in Microbiology in 2004. In the spring of 2006, she entered the Master of Science degree program in Biology at Loyola University Chicago. She currently lives California.